Tomislav Hengl • Michael Gould • Wouter Gerritsma

The Unofficial Guide for Authors

From Research Design to Publication

January 31, 2012

www.Lulu.com

Wageningen Arnhem

Legal Notice: The information presented herein is for informative purposes only and not to receive any commercial benefits. Under no circumstances shall the authors of this Guide be liable for any loss, damage, liability or expense incurred or suffered which is claimed to resulted from use of this Guide, including without limitation, any fault, error, omission, interruption or delay with respect thereto (reliance at User's own risk). This is a self-published document that presents opinions of the authors only and not of the associated organizations and/or user communities. Neither Wageningen University nor any person acting on behalf of Wageningen University is responsible for the use which may be made of this publication.

This is the second, extended edition of the EUR 22191 EN (ISBN: 92-79-01703-9) Scientific and Technical Research series report published by the Office for Official Publications of the European Communities, Luxembourg. This book will be periodically updated. To obtain the most recent version please visit:

http://stores.lulu.com/t_hengl

Contact Addresses:

Tomislav Hengl, ISRIC — World Soil Information, P.O. Box 353, 6700 AJ Wageningen

Tel.: +31-(0)317-484199

E-mail: tom.hengl@wur.nl

http://www.isric.org

Michael Gould Associates BV, Apeldoornseweg 21

Tel.: +31-(0)26 3516750

E-mail: mike@gouldassociates.eu

Wouter Gerritsma, Wageningen UR Library

Tel.: +31-(0)317 483052

E-mail: wouter.gerritsma@wur.nl

http://library.wur.nl

Josip Vranjković, Zagreb, Croatia

E-mail: josip@zanimacije.net

http://zanimacije.net

This document was prepared using LaTeX software. The layout is an adaptation of the Springer Verlag's svmono document class used for monographs. Printed copies of this book can be ordered via Lulu.com. Illustrations by Josip Vranjković (zanimacije.net), Zagreb, Croatia.

ISBN 978-90-817741-0-9

© 2011 Tomislav Hengl, Michael Gould & Wouter Gerritsma

The content in this book is licensed under a Creative Commons Attribution-Noncommercial-No Derivative Works 3.0 license. This means that you are free to copy, distribute and transmit the work, as long as you attribute the work in the manner specified by the author. You may not use this work for commercial purposes, alter, transform, or build upon this work without an agreement with the authors of this book. For more information see http://creativecommons.org/licenses/by-nc-nd/3.0/.

Foreword

When many books offer advice on writing science papers, one could question the need for another. Similarly, when some members of the European Association of Science Editors (EASE), led by Sylwia Ufnalska, suggested that EASE publish some simple guidelines for authors, I queried whether more guidelines were required by the scientific community. But as the worlds of science and publishing move forward, textbooks on writing become dated, or perceived as such by today's new authors, or simply fade from general awareness. Thus, I was pleasantly surprised by the success of the EASE Guidelines for Authors and Translators of Scientific Articles to be Published in English. In the same vein, I welcome this initiative from Tomislav Hengl, Mike Gould and Wouter Gerritsma. The exciting feature shared by both projects is that the material is freely available via the Internet.

The Unofficial Guide has made additional use of the Internet during the writing process, using a wiki-type system that enables multiple authors to read and comment on text at the same time. Such technology will transform the way that students and junior scientists write papers in the future. But, the novice author still has to put finger to keyboard and type those first paragraphs. A daunting prospect, particularly for those who are not used to writing in English. Which is where The Unofficial Guide comes in.

This book offers insight into the world of science publishing that those in large labs, who are producing papers every month may take for granted. It starts with a brief overview of the world of science: again, many will have learned some of this *'by osmosis'* during their science training. Those lucky enough to attend a science conference early in their PhD career will hear new work presented and challenged, but may also hear talk in the bar about rejected papers, the burden of peer review or just the demands of the reviewers for *'one more experiment'* to flesh out the results in what the author thought was a perfectly good paper. For those who haven't, the Guide looks at the reasons for writing research papers, the types of paper and the process of peer review. There is also information on impact factors and Bibliometric services, which is of less direct use to novice authors, but could be useful in their future careers.

The second chapter — on Scientific Publishing — provides more background and the danger is that the author who should be starting to write their own will read this during that prevarication phase. But once it's read, there is no further reason for delay and the meat of the book, The Guide for Authors, will help you through the process. The book's authors rightly emphasize the importance of design and preparation: if your science project was not well designed, you won't be able to put that right at the writing stage, so junior scientists and students should read this part before they even embark on their research project. The Guide offers various rules and tips from the literature, all written in a lively way that make entertaining reading, while presenting solid advice.

Chapter 4 most closely resembles the traditional book on science writing, taking the author through the different steps. The final chapter on submitting the paper also covers traditional ground, such as finding the right journal (something that should be considered at the start of the writing process, not the end) but also addresses matters that have arisen during the 21st century, such as copyright transfer and online availability (which feels very 20th century to some disciplines but it's surprising how many journals are still not available digitally).

I congratulate the authors on producing a practical resource that is fun to read and thus should be not only opened but actually read by many aspiring to join the hallowed ranks of *'published scientist'* so that they may say to their colleagues as someone said to me at my first conference, *"See you in the literature"*.

Joan Marsh MA PhD
President of the European
Association of Science Editors (EASE)

From the authors

Most scientific journals provide guidelines for authors — how to format references and prepare artwork, how many copies of the paper to submit and to which address, etc. Most official guidelines say little about how you should design and produce your paper and what are the chances that it will be accepted. You will also not find such information on journal websites.

This gave us the idea to write an unofficial guide for authors, in which we could tell you frankly what you can expect from journals, editors, reviewers and, indeed, the whole system of science. We offer some practical tips on how to manage the production of your paper. We also address some of the deeper aspects of preparing and publishing research articles as well as the limitations and frustrations that are inherent in current editorial systems such as hyperproduction, gift authors and poor reviews. We then provide instructions on how to improve writing especially at the *meso* (paragraph; logical moves) and the *micro* (sentence) level.

This guide is primarily intended for inexperienced researchers, although we hope more experienced authors will also find some of the points raised in it of interest.

It is clear that, in this book, we strongly promote the Open Source software initiative. Evolution of the internet and Free and Open Source Software (FOSS) have opened up opportunities for the democratization of science and have provided a platform for cross-national and cross-discipline collaboration. By using FOSS, anyone has an opportunity to produce high quality research — human creativity and hard work should be the only criteria. Likewise, internet has made the process of scientific publishing faster, broader and more transparent. Any research organization and any researcher can now be successfully evaluated using public web services such as Web of Knowledge, SCOPUS and Google Scholar. Objective measures are available that can be used to identify the most influential authors, publications and research institutes and distinguish them from work by hyper-productive authors that nobody reads.

Although in this book we mainly emphasize the role of Open Access publishers and FOSS, commercial companies will remain an important (domi-

nant?) part of the scientific publishing business for years to come. Our intention was not to criticize nor to suggest that commercial publishers have exploited researchers, but rather to point to new business models.

Every effort has been made to trace copyright holders of the materials used in this book. The authors apologize for any unintentional omissions and will be pleased to add acknowledgments in future editions.

Wageningen / Arnhem,
August 2011

T. Hengl, *M. Gould*
& W. Gerritsma

Disclaimer

Neither Wageningen University nor any agency thereof, nor any of its employees, makes any warranty, expressed or implied, nor assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed in this book, nor represents that its use would not infringe privately owned rights.

Reference therein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by Wageningen University or any agency thereof. Although much of the information published in this book is used by Wageningen University, no warranty, expressed or implied, is made by Wageningen University as to the accuracy of the data and related materials and (or) the functioning of the software. The act of distribution shall not constitute any such warranty, and no responsibility is assumed by ISRIC as to the use of data, software, or related materials.

This book provides only general guidelines for the design and/or production of research publications. It should not be be used as a reference to any specific journal or publisher, nor should it be associated with any publishing model or company. The authors cannot guarantee that, by closely following these guidelines, you will succeed in getting published (no warranty). When submitting an article to a journal, we advise you to study the official author guidelines.

Acknowledgements

The authors are grateful to a number of colleagues:

David Alexander (Ways with Words) for examples of students'work.

Prof. **Mike Fullen** (University of Wolverhampton) for his kind review of the first edition of the book (and for making us realize that we can do better).

James Hartley (University of Keele), and **Joy Burrough-Boenisch** for detailed comments and suggestions.

Ed Hull (Professional English) for his recipe for scientific articles.

John Maindonald (Australian National University) for pointing us to the right literature.

Alex B. McBratney (University of Sydney) and **Edzer Pebesma** (University of Münster) for their inspiration.

David G. Rossiter (ITC) and **Bob MacMillan** (ISRIC) for their comments and suggestions on a draft version of the manuscript.

To the numerous authors and researchers who we have cited so frequently.

Prem Bindraban (Director of ISRIC) for his support and enthusiasm.

“We’re on a mission from God.”

The Blues Brothers

Contents

Part I The system of science

1 A brief guide to the system of science 3
1.1 Basic principles 3
1.2 Scientific information 9
1.3 Why do we write research articles? 12
1.4 Research quality 13
1.5 Bibliometric services 18
1.6 From draft manuscript to peer review 24
1.7 Types of articles 26
1.8 Types of journals 29

2 Scientific publishing: myths, ideals and realities 33
2.1 Problem areas 34
2.2 Missing reviewers 37
2.3 Fashionable pliability 41
2.4 Publishing companies and models 43
2.5 Publishing efficiency 50

Part II Guide for authors

3 Producing scientific information 55
3.1 Design and preparation 56
3.2 The one-page concept paper 58
3.3 Review your results and repeat the experiment 60
3.4 Investigate your audience 61
3.5 Ten simple rules 62
3.6 Coping with stress 66

4 Writing research articles 71
4.1 Establishing a framework for research papers: a recipe 72
4.2 Basic principles of logical reasoning 74

4.3 Figures and tables (generating and editing graphics) 77
4.4 Citing sources 81
4.5 The art of line editing 85
4.6 Common language mistakes 90
4.7 Can you please be more specific? 92
4.8 On styles 94
4.9 Collaborative writing 96
4.10 Free and Open Source Software 100
4.11 The abstract 103
4.12 The title 106
4.13 Final tips 110

5 Submitting your article 113
5.1 Find the right journal 113
5.2 Do the work of the reviewers yourself 115
5.3 Make a full information package 120
5.4 Pre-submission checklists 121
5.5 The review process and response to reviews 122
5.6 PhD study: a leap in the dark 124

Appendix 129

References 143
Cited sources: 143

Index 147

Part I
The system of science

Chapter 1
A brief guide to the system of science

The main objective of this book is to help you produce better papers, hence it is primarily a scientific writing manual. However, before getting into tips 'n tricks for getting your paper published, we need to go way back before the submission of the paper to address some philosophical aspects of preparing and publishing research articles — the big picture.

It might surprise you that we start by talking about first principles, but less experienced researchers in particular often try to publish their work without having any idea of what science is about: what the basic rules are, where they see themselves in the world of science, and for whom their work is intended in the first place. We will try to answer some of these questions and then use this knowledge to build a foundation for offering more specific advice about producing scientific publications.

1.1 Basic principles

> *"The upshot of all this is that we live in a universe whose age we can't quite compute, surrounded by stars whose distances from us and each other we don't altogether know, filled with matter we can't identify, operating in conformance with physical laws whose properties we don't truly understand."*[1]

Without being too philosophical, here are the essential concepts underlying the system of science:

What is science?

The most concise definition of science is that it is a collection of objective knowledge derived through systematic investigations that are well described and can be repeated[2]. Science is an ever-growing evolution of human knowl-

[1] Bill Bryson in *"A Short History of Nearly Everything"*.

[2] We would like to emphasize three key words from this definition, which we will come back to again and again: objective, systematic and repeatable.

edge about our surroundings and ourselves. This knowledge is often rather soft and needs to be rigorously questioned over and over again. Daniel Boorstin: *"the greatest obstacle to progress in science is the illusion of knowledge"*. What we take as fact is often just a hypothesis with a certain amount of evidence, but we need not get into such debates. What is important is that science is not only an *encyclopedia galactica*, i.e. a systematized record of existing facts. Scientists also contribute new, experimental systems, which may not have any immediate application.

What are the (universal) rules of science?

Although there are no official laws of science, science does have some basic (unwritten) rules — conventions that have evolved over the past two or three centuries. The number one rule of science is that scientific knowledge needs to be built on well-developed arguments and proofs and not on beliefs, opinions or authority. Another important rule is that a researcher needs to present models with a best fit to reality and not personal (subjective) aspirations. For science, even the most pessimistic truth is better than fiction. Another interesting rule is that there is no democracy in science: all can be wrong and a single person can be right as long as he/she can prove it, i.e. as long as he/she can falsify an established proof[3]. Researchers who do not accept arguments, but follow the herd, often discover that the whole community was mistaken. On the other hand, if the existing information is *'soft'*, hence inherently uncertain, pluralism and open discussion, even speculation, must be allowed[4]. Of course, as long as they are backed up by data and strong arguments, science should not rule out any potential solutions and interpretations of unexplained phenomena: everything is *a priori* possible; the issue is only how probable it is. Of course, we are mostly interested in reporting on things that are highly probable.

Another important rule of science is that scientific proofs need to be built primarily through systematic investigations — i.e. research experiments[5]. In addition, the results of research experiments need to be reported in an unbiased, clear, concrete, coherent and logically structured way, again giving much more emphasis to arguments and proofs than to personal feelings[6]. Many people believe that researchers also need to be able to report on new knowledge in an open-minded way, i.e. you should also look beyond your particular scientific *tribe* and attempting to reach members of other tribes of science.

---

[3] *"Truth"* in science exists only until falsified (according to the Popperian approach).

[4] For example, the issue of global warming is still based on limited data; some argue that there is not yet enough evidence to classify it as self-evident. It is therefore advisable to use a variety of models to explain this phenomena and then carefully evaluate them.

[5] Experiments are tests that are systematically designed and described in (more or less) controlled conditions. These can be physical or virtual (and even mind) experiments, i.e. simulations using synthetic data.

[6] This does not necessarily mean that scientists are not passionate people! Blaise Pascal: *"Clarity of mind means clarity of passion, too."*

The last rule worth mentioning is that science does not have a final goal nor final theories[7]. Once a mystery is solved, the focus of science slowly moves to another one. There are certainly no limits to imagination or perfectionism. So, if you think that your word in science will be final or that you will discover the ultimate theory, you might be disappointed. Here are three classical quotes that support this position. FEYNMAN (1965): *"The ideas degenerate, just like the degeneration that great explorers feel is occurring when tourists begin moving in on a new territory."* Albert Einstein: *"The important thing is not to stop questioning; curiosity has its own reason for existing."*[8] John Maddox: *"The big surprises will be the answers to questions that we are not yet smart enough to ask. The scientific enterprise is an unfinished project and will remain so for the rest of time."*

Rules of science[9]:

(1) Scientific knowledge needs to be built on arguments and proofs — not on beliefs or authority.
(2) Scientific (logical) proofs need to be built through research experiments that are repeatable and unambiguous.
(3) The results of research experiments need to be reported in an unbiased, clear, concrete, coherent and logically-structured way.
(4) Logical reasoning (logic) is the only essential skill. Specific laws can always be derived from general laws and regularities.
(5) Even the most pessimistic truth is more valuable than fiction.
(6) There is no democracy in science: all can be wrong and a single person can be right.
(7) Everything is possible *a priori*; the issue is only how probable it is. Proofs only indicate which of the claims is the most likely to be true.
(8) Science evolves through open debate. Pluralism and open discussion, even speculation, must be allowed. Act in good faith and assume good faith from others.
(9) Claims that are purely based on speculation should be avoided.
(10) Researchers should be able to report on new knowledge in an open-minded way.
(11) Researchers need to follow some basic ethical principles in their work: they should avoid fabrication of results, as well as situations where there might be a conflict of interest.
(12) Researchers need to publish their work in publicly accessible (and indexed) media. Society evaluates their work and acknowledges the authors and institutions involved.
(13) New scientific discoveries and theories are associated with institutions/researchers who publish them first.
(14) Scientific knowledge published in research publications needs to be attributed in the manner specified by the author or licensor. Respect copyright laws / avoid plagiarism.
(15) Science has neither a final goal nor final theories.

What are the goals of science?

It depends who you ask: authors, journal editors, and readers have different goals. However, the primary goal of any scientific work should be to make discoveries and explain them. These discoveries/explanations are then used to

[7] There is a whole book on this topic. See WEINBERG (1993).

[8] Albert Einstein as quoted in LIFE magazine (2 May 1955).

[9] Do not try to find these on the internet — we made them up!

solve important problems that allow people to benefit more from their lives. Although many distinguish pure experimental research from applied research, all results, whether experimental or applied, are used for the benefit of other researchers and, ultimately, the wider community. In this sense, science is always application oriented; the only issue is whether users will benefit immediately or later. On the other hand, researchers working on theoretical or experimental topics should not be expected to have to justify their work in terms of direct benefits in the short run. Scientists need a certain degree of creative freedom and political and economic independence. Just like, for example, artists.

Who are scientists/researchers?

Scientists are people who actively conduct experiments and investigations in order to explain phenomena or suggest new ways to improve current systems. The essence of the scientific mind set is: embrace doubt while walking along the winding path toward clarity.

Although scientists are like anyone else, with many different characters and habits, they also have some specific characteristics that differentiate them from others. CREEDY (2008) calls these *"the three Cs of research"*: curiosity, concentration and confidence. In our view, scientists are driven to do science by the following three psychological phenomena: *curiosity*, *imagination* and *perfectionism*. We would also like to add to this trinity *stamina* and *enthusiasm*. As Albert Einstein puts it: *"I have no special talents. I am only passionately curious"*... *"One cannot help but be in awe when contemplating the mysteries of eternity, of life, of the marvelous structure of reality. It is enough if one tries merely to comprehend a little of the mystery every day...."*[10]

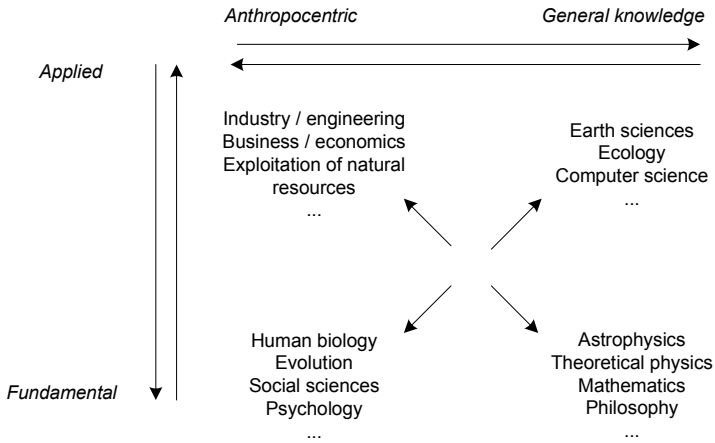


Fig. 1.1 One possible 2D high-order classification of science.

[10] Albert Einstein in a letter to Carl Seelig (11 March 1952).

A broader view on science is that it is a complete human system consisting of *scientific products* (data, information, knowledge, laws, theorems, algorithms, and scientific articles and books), *researchers* (people who professionally conduct research experiments), *research groups* (teams of researchers working on similar topics and meeting regularly at scientific meetings and workshops), *legal entities* (research societies, institutes and schools, national and international organizations with their members, structures, formal rules and evaluation criteria), and *commercial companies* that sell or trade in scientific products (publishers and other providers of scientific information).

At the highest level, scientific fields can be classified as a) *formal* or theoretical and b) *applied*, although such distinctions are rather fuzzy. Another major split in science is that between *natural* and *social* or behavioral sciences (see also Fig. 1.1).

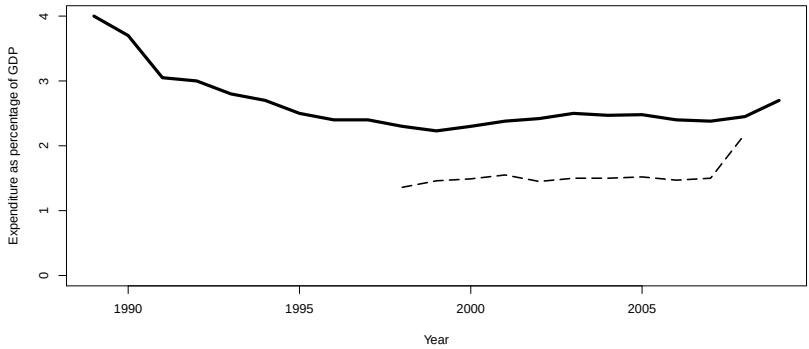


Fig. 1.2 Military (bold line) and research & development (broken line) expenditure as a percentage of GDP (World data). Data source: World Bank, World Development Indicators and `statinfo.biz`.

The contemporary view of science is that the classification of people into (a single) research field is rather old-fashioned and that groups should strengthen their *multi-* and *inter-disciplinarity* through collaboration with other (thematically) distant groups and through combinations of professionals within a single group. The focus in science should be on solving research problems and reporting research in a systematic way and not on formal relationships between and within the fields. Nevertheless, many government organizations insist on research classifications[11]. Standardization of science makes it easier to orga-

[11] For example, the Australian Government (`http://www.abs.gov.au`) requires every research proposal to be linked to a specific research field, discipline, and clear socio-economic objectives, i.e. national research priorities.

nize research applications, find reviewers and link one's work with research priorities.

Not all research fields receive the same degree of support or funding. Unfortunately, many research fields that attract large sums of money are on the edge of being classified as science[12]. Likewise, people spend much more money on producing medication to treat obesity than on research on healthy nutrition or protecting endangered species. We have still not reached a level of civilization where more money is spent on research and development than on weapons (Fig 1.2). Just to give you a rough figure, the Watson Institute[13] has estimated that the costs of the USA military intervention in Iraq and Afghanistan are between 3.2–4 trillion US$ (!). Imagine if this money was spent on research and education instead.

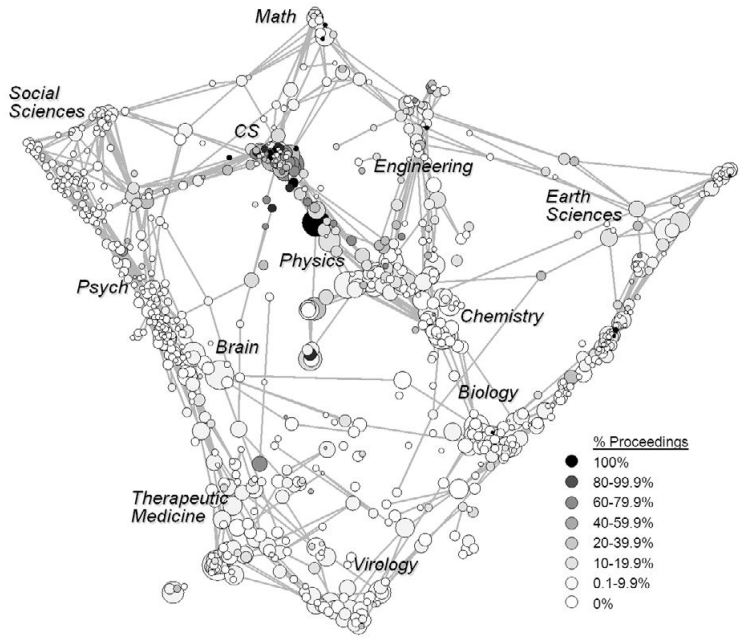


Fig. 1.3 Graphical representation of major scientific fields according to BOYACK (2009): disciplinary map including SCIE, SSCI and Proceedings databases. Each node (circle) is a cluster of journals and is sized to show numbers of papers by cluster. With permission from Springer Science+Business Media.

Should every researcher strive to achieve multidisciplinarity? Researchers live, create and move around within what are often closed circles (research

[12] For example, secret military experiments receive far more funding than public benefit/democratization projects such as the One Laptop Per Child or WIKIPEDIA.

[13] `http://costsofwar.org`

fields), which have their own rules and principles. Isolation of scientific fields in contemporary science is considered to be old-fashioned, but it still happens. For Howard Gardner[14] hyperdisciplinarity or looking at everything through one discipline is an example of a poorly disciplined mind: *"Economists who see the whole world through rational choice; psychologists who see the whole world through evolutionary psychology; the lawyer who sits down with his children who are two and three years old and writes down a constitution which gives the children their rights and their responsibilities."*

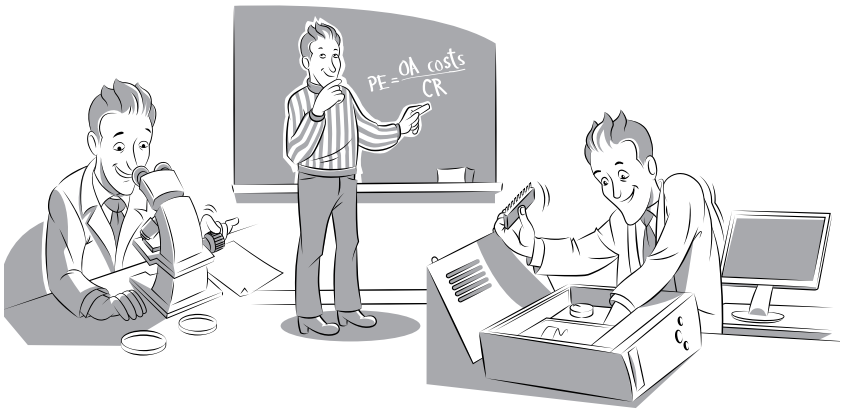


Fig. 1.3 shows a (disciplinary) map of some major scientific fields based on numbers of publications in the Science Citation Index Expanded (SCIE) /, Social Sciences Citation Index (SSCI), and Proceedings databases (BOYACK, 2009). This shows that many research groups are clustered within a field or subfield, which is probably related to the way universities, research institutes and journals are organized — faculties and departments are often isolated; graduates then go on to work in professional associations, which have their own unions and even political lobbies. Each field may have its own principles and people in authority. Fig. 1.3[15] shows the impact of political priorities on science.

1.2 Scientific information

Scientific information is the collective name for any type of written, physical or multimedia record of scientific work. For practical reasons, the products of research work (research projects) are typically split into one or more standalone items (*scientific information packages*) which are used to convey a new

[14] `http://www.howardgardner.com`

[15] There are several similar world maps of science; e.g. by BOLLEN *et al.* (2009).

concept or discovery to a research community (except in the case of a review article, in which authors attempt to systematize existing knowledge). Scientific information packages are usually published in written media — as journal articles, books or book chapters, websites, but also increasingly in multimedia, including videos and interactive websites and programs. Popular explanations of scientific information are then often communicated more widely through television and the print media.

The process of producing and disseminating scientific information is called publishing. Scientific information (packages) are usually published in written media — as journal articles, conference proceedings, technical reports, textbooks and/or book chapters.

Scientific information evolves during the research process. Researchers start with some rough ideas and initial data and then build these up into standardized products (publications, multimedia materials) that are searchable through commercial or open indexing systems. The role of commercial companies is to register, certify, distribute/promote and archive scientific information.

On copyright and authorship — Copyright is the exclusive right granted by the law of a jurisdiction to the product owner to copy, distribute and make profits. Copyright should not be confused with *"authorship"* — the right of a person or group of people who developed an original product (the creators). Authors make intellectual property (IP) that is automatically copyrighted. This may be their own property, or it may belong to the organization that finances their work (it's often specified in employment contracts). A good example: Paul McCartney wrote music for the Beatles and Michael Jackson bought the copyright. Authors can transfer their copyright to individuals or commercial companies. By transferring the copyright, authors no longer *'own'* their intellectual creation and need to request permission to copy and distribute it according to a licence agreement. In most countries copyright ceases after 70 years after the death of the author. Authorship is non-transferable. In many legal systems around the world authorship also includes some rights. For example, a publisher with a copyright on some work has no right to alter, modify or blank out the original authors' names. In most legal systems the main concern is the copyright, which gives its holder the right to financial benefits.

In more general terms, there are at least five forms of scientific information products:

- *Unprocessed products*: raw data, development versions of algorithms, notes and sketches.
- *Draft products* (test beds): development versions of information systems, technical reports, draft documents, beta versions of packages (new software and/or technology).

- *Pre-prints* (no peer review, no editing): submission-ready products, rigourously tested versions of information system, scientific documents (articles) and beta versions of software packages. Publication of pre-prints allows scientists to get new work out quickly, although with limited credibility.
- *Published products* (peer reviewed, edited): licensed (standard) scientific products that are indexed in scientific databases i.e. those with a patent number, Digital Object Identifier (DOI), ISBN or similar.
- *Promotional by-products*: media packages developed to support the promotion of scientific information — slide shows, websites, newspaper articles, documentaries, pubcasts, and videos.

These are the basic evolutionary stages of scientific information. Commercial companies are typically only interested in submission-ready papers that can be prepared and copyrighted as commercial products. Researchers, on the other hand, work with scientific information products at all stages. In this book we present guidelines for how to produce all of the products listed above.

The process of producing and disseminating scientific information is called publishing. There are three major publishing models:

I *Commercial (toll-access) publishing* — The author transfers the copyright to the publisher in exchange for professional distribution. Copyrighted materials are sold in bulk or in parts. Copyright infringement is regulated by law and subject to financial sanctions.

II *Open access (subsidized) publishing* — Authors or their organizations pay[16] publishers to *open* the copyright and make the materials available free online. There are three versions of Open Access (OA) publishing:

a. *Author-pays OA* — the costs of OA are paid by the author.
b. *Organization-pays OA* — the costs of OA are paid by an academic and/or government institution.
c. *Delayed OA* — the article is made available for free download after a fixed period of time, which is usually longer than one year[17].

III *Open (non-commercial) publishing* — All materials (manuscript and the review process) are copyrighted under public domain. These can be freely distributed without limitation.

a. *Organization supported* — production and printing are supported by an organization.
b. *Self-publishing* — production and printing are organized solely by the authors.

[16] In some situations publishers might give up the right to profit and provide open access free of charge.

[17] At an International Federation of Science Editors conference in Merida, Mexico in 2004, a doctor pointed out to an Elsevier executive that *"by the time you make information available at a price I can afford, my patient is dead."*

Note that the annual subscription to a specialist medical journal can cost the price of a small car; author-pays OA can also cost a significant amount of money (see further page 45).

1.3 Why do we write research articles?

Have you ever asked yourself what the essence of a research article is? Is it merely a report on an experiment or is it an essay or a user guide for colleagues who would like to conduct similar research? In fact, a research article is a bit of all of these. It looks like a research report, but it also contains some unique spices: belletristic elements similar to those in an essay or article in a magazine. However, unlike essays, research articles need to follow a logical structure and provide all technical details. They must be rigorous because they have a practical purpose — to communicate the results of research in a systematic and standardized way.

The purpose of a research paper is to communicate the results of research — new, original findings (new concepts, new data) — in a systematic and standardized way so that readers can apply, modify and extend that knowledge.

People have different reasons for publishing their research. First, there is prestige — researchers aim to be the first to explain or solve an important problem. Second, by solving some research problem we get a chance to improve peoples' daily lives *"at scales expanding human lives"* (ASCHERON and KICKUTH, 2004). Third, there is the intellectual exercise, which helps us to express our creativity. Intellectual creations can be intrinsically valuable.

Although researchers also communicate their ideas at conferences, meetings and through the educational system, their contribution to science is mainly measured through research output in publications. Many researchers in the history of science who did not publish their work are no longer connected with that work[18]. This is nicely illustrated by the well-known aphorism *publish or perish*. To some extent, this gives scientific work a competitive character — scientific discoveries are connected with those who first publish them (in a high-impact journal). That's the name of the game.

Publication of your ideas/discoveries is not (at least it should not be) the ultimate ambition of a researcher. The most important thing for a researcher is to make a significant contribution to science, i.e. to make an impact. So the true motto of a researcher should in fact be: *publish and make an impact or perish*.

[18] The classic example of a researcher who *perished* is Christian Huygens, who may well have been a superior scientist to Descartes or even Newton, but did not seriously consider publishing his work (CRUMP, 2002).

1.4 Research quality

By publishing their research, scientists expose themselves to critical evaluation. The scientific quality of a research article is typically a product of: (1) the quality of the design of the experiment, (2) the quality of arguments/proofs, (3) the quality of methods and materials, and (4) the quality of the presentation of the research results (see Fig. 5.3). Scientific quality is difficult to evaluate before publication. Researchers need time (sometimes even decades) to look at someone else's work and find out what its real advantages are. That's probably why the Nobel Prize is awarded to researchers toward the end of their careers (the median age of Physics Nobel Prize recipients is 51).

Retrospectively, scientific quality can be measured as the effect of a number of professional achievements:

- *Proven impact on the world of science* (good citation statistics)
- *Proven general interest in your work* (invitations to give keynote speeches and lectures, number of articles in the media)
- *Grants received and research funding*
- *Professional awards received* (e.g. best paper awards).

Although these are very concrete measures, they are not easy to collect and update. In fact, the managers of most research organizations do not really know what the performance of their employees is, how established the people in their teams are, and what their performance trend is. Evaluating researchers' performance is usually outsourced to scientific information and consulting companies. The most accepted way of tracking authors' impact is through the tracking and analysis of citations.

> **Bibliometrics** — a subfield of scientometrics — is a set of methods that's used to monitor bibliographic information i.e. library items. By monitoring the citation statistics of research publications we can compare the productivity and impact of different authors and institutes and discern trends and clusters. Several bibliometric indices can be used to track researchers and their output. The most widely used are: *h*-index for authors, Impact Factor (IF) for journals, and Citation Rate (CR) and Relative Impact Factor (RIF) for articles. Nowadays, thanks to Google Scholar, SCImago Lab, ResearcherID and other open access web services, anyone can track articles and see which publications, authors and journals are really *'hot'*. For more details about state-of-the-art bibliometric indices see e.g. HARZING (2010).

Once an article is published, it starts accumulating *'points'* i.e. citations. This is the absolute measure of its impact. Of course, it's not fair to compare the number of times a 15-year-old paper has been cited with the impact of the work of a *'rookie'*. Citations need to be normalized e.g. by dividing them by article age and by the research field.

An objective measure of the impact of a paper is the average number of annual citations — its **Citation Rate** (CR) (GARFIELD, 1990):

$$CR = \frac{\text{total number of citations}}{\text{years since publication}} \tag{1.1}$$

CR is commonly filtered for self-citations. A further modification of the CR is citation rate per author (CRpA), which normalizes for the number of authors for each paper (HARZING, 2010).

By dividing the citations by the average citation number in a research field one can derive the **Relative Impact Factor** (RIF):

$$RIF = \frac{\text{total number of citations}}{\text{average number of citations in the field}} \tag{1.2}$$

For example, papers in environment and ecology receive an average of 11.35 citations (Fig. 1.4[19]). This means that a paper that receives 20 citations has an RIF of 1.76.

Most articles gradually disappear from references, which means that they have a citation *'half-life'*: the number of years that one has to go back in time to account for 50% of the total references (GARFIELD, 1990). In Natural Sciences, the citation half-life typically ranges from 3-10 years, although most articles will only be cited during the first few years (if at all).

An objective evaluation criteria of the impact of a research article is its Citation Rate — the number of times it has been cited per year. Citations can be further standardized per research field using global averages to estimate the Relative Impact Factor.

A publication's citations plotted on a time line approximately follow the asymmetric (Hubbert) curve: they will first grow exponentially, reach a peak, and then follow a decay function (Fig. 1.5a). Unlike books, which can be updated periodically in new editions or updated continuously online (like this book), research articles get a permanent bibliographic reference and tend to have a limited life. With exception of Scholarpedia, which promotes the idea of having *live* articles — on-line articles which are continuously updated by article *curators*.

Unfortunately, many publications have almost no life at all i.e. they never get cited. Most researchers will only write a handful of publications that make a real impact in their field. If we sort an author's publications according to the number of citations, we obtain a graph such as that shown in Fig. 1.5(b).

In order to track scientists' output using a single measure, HIRSCH (2005) proposed a simple citation index that was then named "*h*-index" after Hirsch's name. An ***h*-index** is derived from the number of an author's papers that are cited at least *h* times (see Fig. 1.5). The *h*-index is more suitable for evaluating authors because it corrects for *"one-hit wonders"* — academics who have authored only a small number of highly-cited papers (ROEDIGER, 2006; HARZING and VAN DER WAL, 2009; BAR-ILAN, 2008).

[19] Based on `http://researchanalytics.thomsonreuters.com`

Field Rankings

Sorted by: Citations per Paper SORT AGAIN

	View	Field	Papers	Citations	Citations Per Paper
1		MOLECULAR BIOLOGY & GENETICS	282,538	7,237,466	25.62
2		IMMUNOLOGY	125,685	2,741,500	21.81
3		NEUROSCIENCE & BEHAVIOR	303,817	5,916,219	19.47
4		BIOLOGY & BIOCHEMISTRY	569,918	9,829,215	17.25
5		MICROBIOLOGY	166,517	2,629,433	15.79
6		SPACE SCIENCE	125,769	1,799,022	14.30
7		CLINICAL MEDICINE	2,116,193	27,355,596	12.93
8		PHARMACOLOGY & TOXICOLOGY	186,449	2,275,570	12.20
9		ENVIRONMENT/ECOLOGY	267,167	3,032,396	11.35
10		PSYCHIATRY/PSYCHOLOGY	247,522	2,786,509	11.26
11		CHEMISTRY	1,197,020	13,394,419	11.19
12		GEOSCIENCES	280,749	2,722,183	9.70
13		PHYSICS	883,905	7,924,536	8.97
14		PLANT & ANIMAL SCIENCE	560,085	4,337,330	7.74
15		AGRICULTURAL SCIENCES	204,677	1,442,598	7.05
16		MATERIALS SCIENCE	458,967	3,228,240	7.03
17		ECONOMICS & BUSINESS	167,278	1,040,131	6.22
18		MULTIDISCIPLINARY	19,070	95,468	5.01
19		ENGINEERING	817,334	3,887,615	4.76
20		SOCIAL SCIENCES, GENERAL	455,440	2,126,723	4.67
21		COMPUTER SCIENCE	264,026	989,288	3.75
22		MATHEMATICS	263,350	915,152	3.48

Copyright © 2011 The Thomson Corporation

THOMSON

Fig. 1.4 Field rankings table based on Essential Science Indicators data (a bibliometrics product by Thomson Reuters).

An author with a high *h*-index has delivered a durable academic performance. The *h*-index is basically designed to distinguish truly influential scientists from those who simply publish many papers. The problem is that it is a function of the scientific age of an author and it may be much smaller for relatively young authors who might be writing just as good articles as their senior counterparts.

The third common bibliometric measure is the Impact Factor (IF), which is defined as:

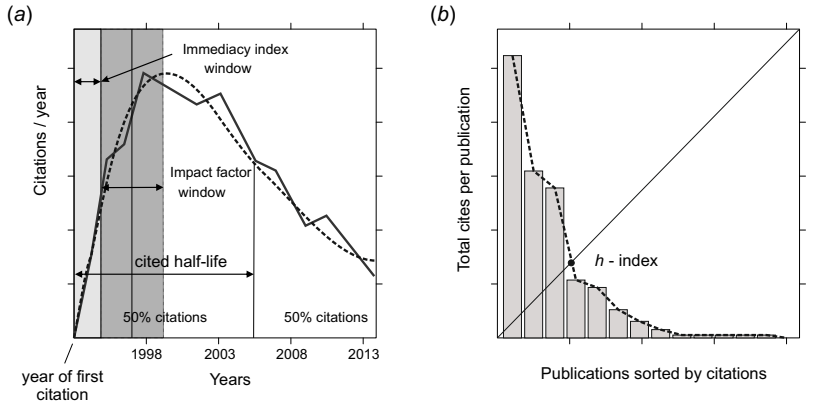


Fig. 1.5 Basic principles of citation analysis: (a) an article's life visualized using citations per year; (b) the derivation of the *h*-index.

$$IF(t) = \frac{\begin{array}{c}\text{number of times articles published in year } t_{-2} \text{ and } t_{-1} \\ \text{which were cited by indexed journals in year } t\end{array}}{\begin{array}{c}\text{total number of citable items} \\ \text{published by that journal in year } t\end{array}} \tag{1.3}$$

IF is a bibliometric measure that was developed primarily for evaluating journals, i.e. to derive a Journal's Impact Factor (JIF). JIF has received a lot of criticism from researchers and research organizations for various reasons. First, often many articles, even in journals with a high impact factor, are almost never cited (SEGLEN, 1997). Research on Nature's 2004 impact factor, for example, has shown that about 90% of its IF was based on only a quarter of its publications (NATURE EDITORIAL, 2005). Second, journal issues are artificial entities. Articles in a journal issue are independent, even when they come in special issues. To mix the achievements of people who have most probably never even heard of each other is like mixing apples and oranges (as the plot in Fig. 1.6 left illustrates). Third, a journal can adopt editorial policies that artificially increase its impact factor. THE PLOS MEDICINE EDITORS (2006) warn that a consequence of basing the evaluation on JIF is that *"science is currently rated by a process that is itself unscientific, subjective, and secretive."*

Unfortunately, JIF is still by far the evaluation criterion that's most widely used by government and funding agencies (ADLER *et al.*, 2009). If you plan to send an application for a scholarship or funding, there's a good chance that evaluation of your proposal will be primarily based on the IF of the journals in which you published your work.

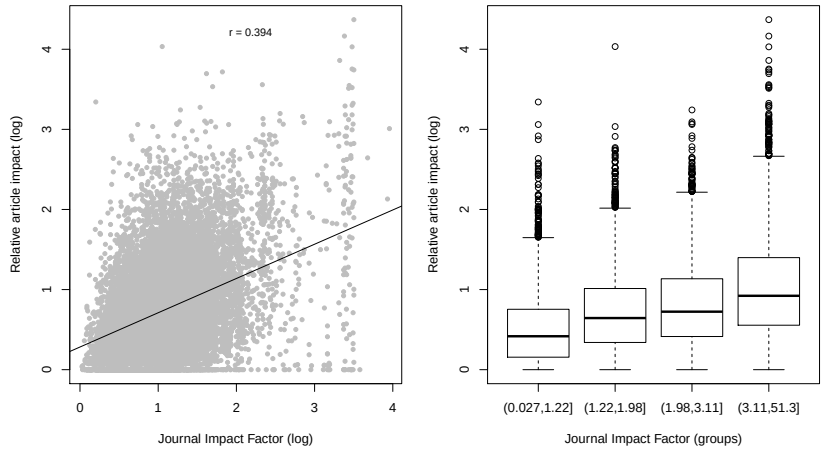


Fig. 1.6 Journal impact factors and individual relative impact factor (Eq.1.2) for 14,348 articles published at Wageningen University in the period 2002-2009. JIF explains only 14% of the variability in the RIF. The original data can be obtained from the Wageningen University Library.

JIF is often incorrectly applied to evaluate the significance of an individual publication or an individual researcher. However, if you submit an application for a scholarship or grant, there's a good chance that the evaluation of your proposal will be primarily based on the JIF related to your publications.

Nevertheless, work by LARIVIÈRE and GINGRAS (2010) has shown that duplicate papers published in higher-impact journals obtain, on average, twice as many citations as their identical counterparts published in lower impact factor journals. Hence high IF journals do make a difference (see also Fig. 1.6 right), and journals with a JIF in the upper quantile do perform better on average. LARIVIÈRE and GINGRAS (2010):

> *"The intrinsic value of a paper is thus not the only reason a given paper gets cited or not, there is a specific Matthew Effect attached to journals and this gives to papers published there an added value over and above their intrinsic quality."*

In other words: publishing in high-impact journals is good, but using JIF as a measure of the quality of individual articles is controversial.

The good news is that bibliometrics are slowly changing towards more diverse, more web-based measures. PLoS has recently introduced *"article-level metrics"* — a list of measures that focus on individual merits, rather than on the journal's impact factor. These include[20]:

[20] `http://www.plosone.org/static/almInfo.action`

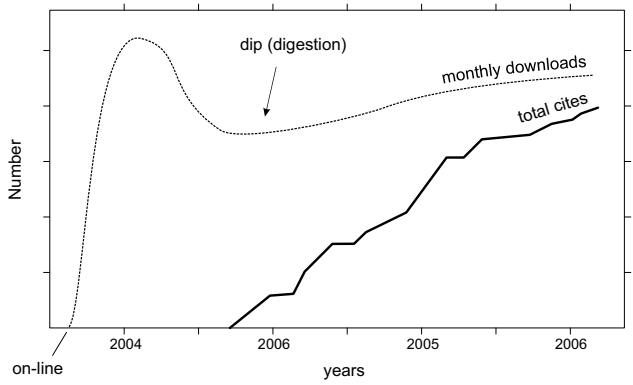


Fig. 1.7 Schematic example of download (or paper access) statistics versus citation statistics. Although download statistics and citations are likely to be correlated, citations are typically delayed for 2-3 years; likewise, the most downloaded papers do not necessarily influence citations.

- Citation statistics by third-party citation measuring services (Scopus, PubMed Central, and CrossRef)
- Number of article views / visitor statistics (PDF, HTML, XML format of documents)
- Social bookmarks at Delicious, CiteULike and Connotea
- Comments and notes, blog activity, article rankings and other similar types of web activity.

Web-based measures will play an increasingly important role in bibliometrics.

1.5 Bibliometric services

Nowadays the impact and importance of research publications, journals, and authors can be successfully followed through web services, the three best known of which are:

- **Web of Knowledge** (`webofknowledge.com`) — is a Thomson Reuters' Scientific subscription-based multidisciplinary database that covers about 9000 peer-reviewed journals clustered in the following three databases: the Science Citation Index Expanded, Social Sciences Citation Index Expanded and Arts and Humanities Citation Index. The v4.0 of Web of Knowledge indicates that it contains about 23 million full documents from the period 1988–2011. Web of Knowledge can generate citation reports and be used to analyze the results, e.g. to compare the success of authors, institutes or

countries, and to see how paper citations change over time. The results of searches can be exported in a variety of formats and used to generate reports. Web of Knowledge is the most detailed and most accurate scientific database of peer-reviewed articles published in the English language (but available only by subscription).

- **SCOPUS** (`scopus.com`) — Elsevier's SCOPUS (also subscription-based service) contains around 33 million abstracts from about 16,000 peer-reviewed journals, but also shows the status of non-registered SCOPUS publications, including 386 million web-based publications. SCOPUS has recently offered a Citation Tracker service that makes it possible to assess impact of an individual author/publication. The *h*-index is now also incorporated in SCOPUS and can be derived for each selected author/group. A limitation of SCOPUS is that it does not contain citation information for articles published before 1996.
- **Google Scholar** (`scholar.google.com`) — GS is a non-commercial academic search service that registers publications available on the web. It indexes all on-line materials, including PowerPoint presentations, mailing lists, blogs and such-like, but also all peer-reviewed journals that are available on-line (except those published by Elsevier). The advantage of Google Scholar is that it allows free searches for publications written in any language and from any publisher and thus contributes to the democratization of citation analysis (HARZING and VAN DER WAL, 2008; HARZING, 2010). Google Scholar has been available since 2004 and its quality is continually improving. Its biggest limitations are noisy data, inconsistencies, duplicate publications, and errors. Microsoft now also provides an academic search service called **Microsoft Academic Research**[21].

Each of these web services has its advantages and disadvantages and each offers something that its competitors do not have. They clearly compete with each other in providing key information about citation records and there will always be differences — some minor, some significant.

BAUER and BAKKALBASI (2005), for example, showed that there is not much difference between GS and Web of Knowledge in terms of the accuracy of assessing the number of citations for highly-cited publications, but there is indeed a significant difference between GS and Web of Knowledge / SCOPUS in assessing all publications within a certain field. MEHO and YANG (2007) estimated that the overlap between the results of GS and Web of Knowledge / SCOPUS is only 30–50%. This happens typically because GS indexes about twice as many publications as Web of Knowledge / SCOPUS, including conference papers, dissertations, theses, and book chapters. As Google's database grows, this difference is becoming smaller and smaller (HARZING, 2010).

Although there has been pressure for academic organizations to start using GS, many librarians advise sticking to Web of Knowledge or SCOPUS if you need an accurate citation count (GILES, 2005). This is mainly because GS is

[21] `http://academic.research.microsoft.com`

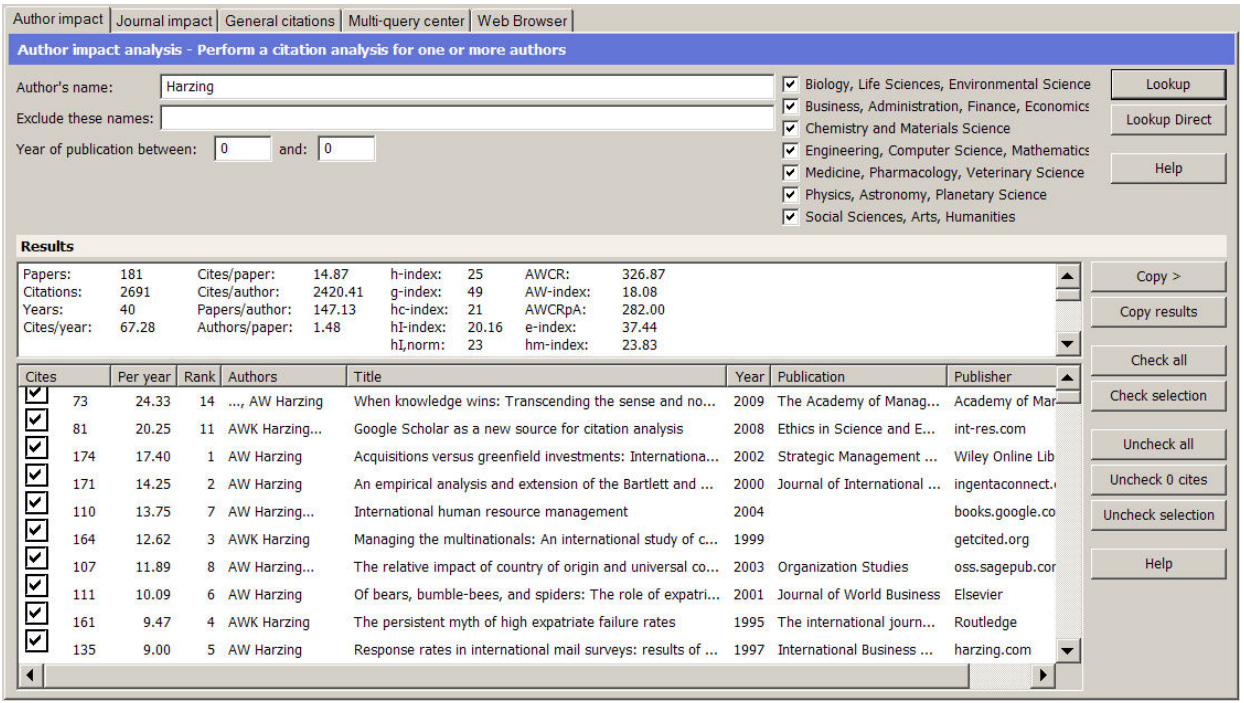


Fig. 1.8 Publish or Perish is a small freeware application, that allows searches of the Google Scholar database and citation analysis (HARZING, 2010). A similar ranking service for journals is the SCImago Journal and Country rank service.

often incomplete and noisy. In addition, Google refuses to reveal details of its search algorithm, although this may change in the near future.

Most international publications can also be browsed using the **WorldCat** service[22], which is provided by the OCLC Online Computer Library Center, Inc. This web service contains over 1 billion items from more than 10,000 libraries. It allows you to search bibliographic items by title, subject and/or author's name. The query results are grouped by authors, research fields, formats, languages and year of publication. WorldCat can also sort query results by publication date and relevance. Each book, article, report, audio or video material receives a unique OCLC number. However, it does not (yet) provide any bibliometric evaluation of publications.

The world's largest internet bookshop **Amazon** also does not provide any scientometric measures. However, it does allow sorting of books by popularity (*sort by best selling*), which can be as important as the most sophisticated bibliometric index.

Nowadays, scientists can also be traced by geographical location. Springer hosts a service called AuthorMapper[23] that shows a geographical distribution of authors and their work (actually it is a space-time visualization) — an excellent way to find out where the scientific hot-spots are. Unfortunately, as with many commercial companies, the results of searches are typically limited to Springer-connected or Springer-owned products.

The main problem of bibliometrics is that scientists do not have a unique international identifier, something like an ISBN for books. At the national level, certain countries have organized some kind of registry, at least for people working for government organizations. These registries are not compatible at the international level, which is clearly inefficient. Many authors have the same names (even with two middle letters), many change locations, some change names, etc.

To account for this, SCOPUS uses an **Author Identifier**, but this is then linked to the hosting institution. For example, the first author of this book has several identifiers:

```
AU-ID("Hengl, Tomislav" 6602864740)
OR AU-ID("Hengl, Tomislav" 13409408400)
```

where `ID 6602864740` refers to the period when the author worked at the ITC in Enschede, the Netherlands and `ID 13409408400` to the period when he was an employee of the JRC at Ispra in Italy. Furthermore, SCOPUS allows users to group the same authors with multiple identifiers. Web of Knowledge does not make this distinction — it only keeps the most recent address in the database. On the other hand, an advantage of Web of Knowledge is that it allows users to zoom into research articles and observe how citations change for each article.

Especially people with *'van'* and *'der'* prefixes are in a difficult situation because they could lose out in the citation rankings just because they are cited in-

[22] `http://worldcat.org`

[23] `http://www.authormapper.com`

consistently in different publications. For Jackie Senior (a science editor at the UMC in Groningen), the worst example so far is certain Dr. Johannes Kristian Ploos van Amstel, also know as Hans Kristian P van Amstel, HK van Amstel, J van Amstel, and many other combinations. Performing a complete bibliometric analysis with such inconsistent author names is very cumbersome. The name ambiguity problem is probably one of the most serious problems of bibliometrics at the moment.

Another example of a global registry of researchers is **ResearcherID**[24] from Thomson Reuters (free service). In August 2011 this contained records from over 100,000 scientists around the globe. ResearchID provides a unique ID for each researcher, which is available via a URL, for example:

`http://www.researcherid.com/rid/F-7147-2010`

The ResearcherID entries are further linked to ISI publications, which are sorted by total number of citations (Fig. 1.9). The problem with ResearcherID is that authors are responsible for setting up and updating their profile, so the total number of researchers in the system is still low.

Like Google Scholar, Microsoft Academic Research also crawls the web and generates researchers' profiles in a semi-automated manner. Microsoft uses a 7-digit number:

`http://academic.research.microsoft.com/Author/1876597`

Unlike with SCOPUS, information about researchers from Google Scholar and Microsoft Academic Research (including citation statistics) is publicly available and hence can easily be queried from the web. Moreover, Google and Microsoft allow anyone to edit the author details to ensure accurate data.

To minimize duplication and errors, unique identifiers such as Digital Object Identifier (DOI) and ISBN for books will increasingly be used. Similar universal identifiers probably need to be introduced for authors and other library items, such as computer programs and maps.

It would be great to have a unique web-based database, which anyone could use to check the correct reference of all library items in the world at any time. Open and efficient bibliometrics and scientometrics registries would certainly contribute to the democratization of science and weaken the monopolistic positions of some scientific information indexing companies. Although Google seems to be closest to reaching this goal, probably some truly non-profit international association such as the Online Computer Library Center[25] would be a better choice.

Another non-profit organization set up to solve the problem of author name ambiguity is ORCID Inc[26]. ORCID is building a global database of researcher

[24] `http://www.researcherid.com`

[25] `http://www.oclc.org`

[26] `http://www.orcid.org`

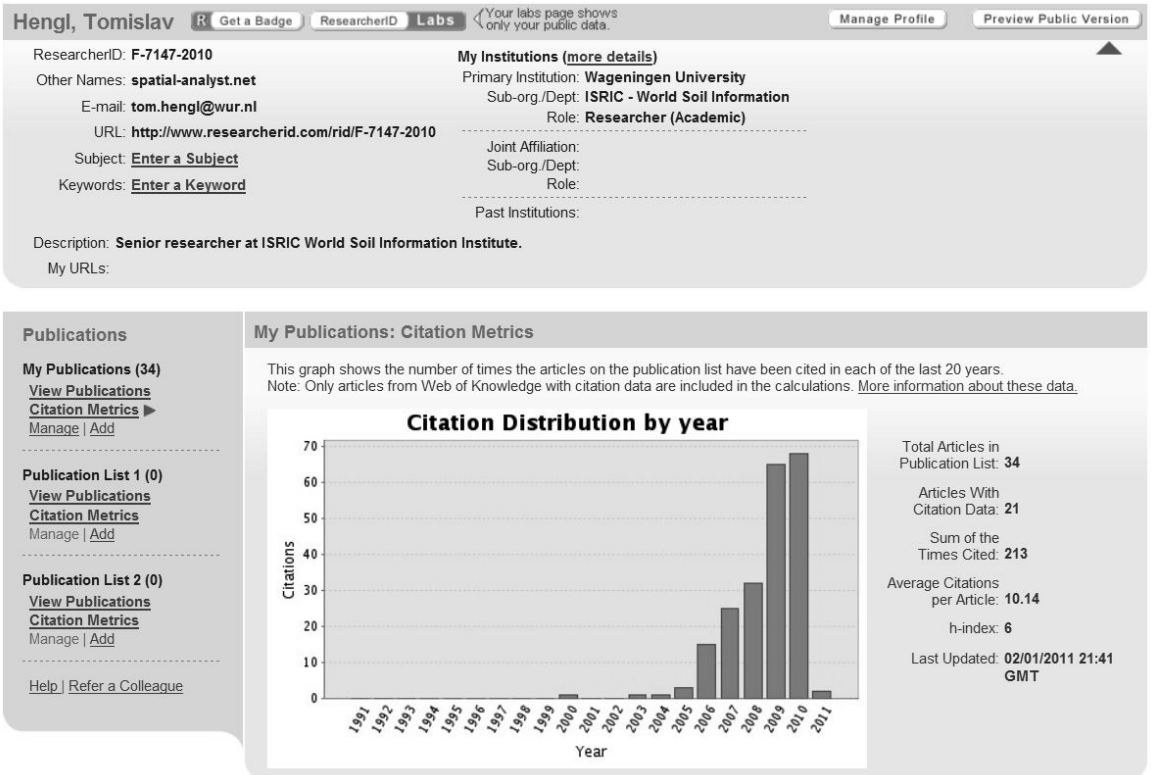


Fig. 1.9 Example of complete citation statistics as reported by ResearcherID.

IDs, which link each author to their IDs in SCOPUS, Web of Knowledge, ResearcherID and similar author databases. Ultimately, we only need one ID in science.

Examples of web services for maintaining and/or cross-linking bibliographic records are **PublicationsList**[27] and CrossRef[28]. Such services allow researchers and research organizations to maintain a reliable web-based record of their academic output. If you already have some publications, you should register them on ResearchID.com and/or PublicationsList.org and thus contribute to building a global registry of authors and publications.

1.6 From draft manuscript to peer review

The process of getting your work evaluated is called **peer review** or refereeing. Peer review is a critical examination of your work by established researchers in the field — usually the journal's reviewers. Based on the degree of anonymity, reviews can be divided in:

1. *Blind* — authors do not see the names of referees (i.e. anonymous reviews).
2. *Double-blind* — authors' and referees' names are hidden.
3. *Personalized* — authors' and referees' names are known.

For most editorial offices, the results of reviews are often left unpublished — stored in an internal database or recycled. There is much debate in the academic world about *'open'* and *'closed'* refereeing and publishing models. For example, Prof. James Hartley, Research Professor in the School of Psychology at the University of Keele (widely known for his work on student learning, text design and academic writing), believes that, especially in the information age, reviewing should be open i.e. the authors should know the names of referees and *vice versa*: *"little — if anything — should be hidden from different contributors to the total system"*. Read more about such debates on page 43.

Review is a filtering process: experienced established researchers help you improve your paper so that it meets certain quality criteria. Filtering scientific publications involves the following specialized operations:

Checking for general suitability
: Journal or book editors quickly browse your document and check if it is of interest to the journal, submitted in the required style and structure, and contains no nonsense. This type of gross-error filtering generally takes very little time. You could compare it with the spam filter in your e-mail manager. However, it's crucial that the editor gets the point of your research very quickly. According to Professor Edwin Gale[29]: *"I receive about 15*

[27] `http://publicationslist.org`

[28] `http://www.crossref.org`

[29] Editor of Diabetologia, as reported by Ed Hull in a workshop on Advanced Science Editing on February 18, 2011.

articles per day. Most of them I reject within 3 minutes. The main reason for rejecting them is that I do not see the point of the research." Similarly, John R. Benfield wrote in the Journal of Medical English Education: *"Thirty years of service as an editor and reviewer have taught me that imperfections or errors in grammar are rarely, if ever, responsible for the rejection of a manuscript. Manuscripts are rejected because they lack new information, new ideas, clarity and credibility."*

Cross-checking the scientific content

In the next phase, journal editors invite (at least) 2-3 experienced scientists who specialize in a similar field or sub-field to focus on the scientific content more closely and evaluate the originality and quality of your methods and writing. This is probably feasible within several days, but because reviewers do this work voluntarily, they are usually given a few months to respond.

Cross-checking the data analysis and results

Ideally, each paper submitted for publication that contains some type of statistical analysis or summaries of results should be checked for accuracy and typos in the code. Recall the Rules of Science from page 5: anybody should be able to reproduce your results. This is not considered to be the responsibility of a journal — publishers make it clear that they are not responsible for the accuracy of numbers or graphs. In fact, data analysis steps are often kept secret, both for objective and subjective reasons. For example, many large software companies keep their code and data *closed* for commercial reasons i.e. to protect their copyright.

English language editing

The communication quality of text can usually be considerably improved. An English language editor does not necessarily need to be an expert in the field, although it is much more efficient if he/she has at least some general knowledge of the topic. Language experts can edit text endlessly, but our concern is not to have a perfectly written paper — simply to have a document that is readable, clear, concise, well-organized and credible[30]. Journals and book publishers rarely invest in language editing. In more than 90% of cases readability is left to the authors.

Editing graphics

A figure in a research article can potentially become better known than the article itself. Charts not only inform, but also can be used to persuade and even campaign[31]. Producing artwork is costly, so it is (unfortunately) often left to the authors, who usually lack design skills (see further section 4.3; page 77).

[30] British and USA embassies organize the certified English language exams IELTS and TOEFL, which are often used by universities as admission criteria; The EU has published a Common European Framework of Reference (CEFR). An author submitting a paper to a journal should pass the *"proficient users"* of English level (at C2 on the CEFR Global Scale) or better.

[31] `http://www.economist.com/node/10278643`

Page proofs

Final polishing of the document — prepared as *Camera Ready Copy*[32] (CRC) — is often referred to as correcting the *"Page proofs"* or proofchecking. Now the author has very little time (24 to 48 hours) to read the document one more time line by line and make corrections. The focus at this stage is on typos, wrongly ordered tables and graphs and minor errors in the text.

As reported in a December 2010 article in Nature, entitled *"A helping hand"*, Jim Viccaro, editor of the Journal of Applied Physics, routinely recommends editing services to authors: *"Reviewers these days are overburdened, and a properly written paper is just easier to review,"* he says. Prices, which vary according to the level of service, the length of the paper and the turnaround time, can be anything from 250 US$ for a 6,000-word paper with a 14 to 21–day turnaround to 5000 US$ for a 12,000-word paper with a 48-hour turnaround.

Viccaro sees it is a worthwhile investment: *"This is how an author can make sure their paper is not dismissed for the wrong reason, just because no one could understand what they were talking about,"* he says. In the same article, Xiao-Fan Wang, an associate editor at the Journal of Biological Chemistry, maintains that directing non-English-speaking authors to editing companies before submission has allowed him to accept an extra 5 to 10 percent of papers that he would otherwise have rejected. *"These services can offer a lot of value,"* he says. *"Not only in terms of the English, but in highlighting what the author didn't even realize was the most important part."*

So to get a document from *Limbo* to *Purgatorio* and on to divine purity often requires expertise from a number of professionals. Unfortunately, researchers often do not have graphic designers, language experts, marketing or IT gurus in their teams. You should at least be aware of what's needed to deliver a polished product and, where necessary, bring in the missing expertise.

1.7 Types of articles

Generally speaking, there are three main types of published scientific articles:

- *Original research articles*
- *Review articles*
- *Popular articles* (or research articles adjusted to target a specific audience)

In this book we focus on designing and producing original research articles, i.e. complete technical information packages representing original new findings. Writing review articles and popular science articles requires different skills from those needed to produce original research articles (TURABIAN, 2007). MONTERO and LEÓN (2007) go further and suggest a classification system for studies in psychology, with three main groups:

[32] Although an outdated term, CRC is still used for digital documents that are ready for printing or press-ready.

I Theoretical studies:

a. Classical reviews
b. Meta-analysis

II Empirical quantitative studies:

a. Observational descriptive studies
b. Survey descriptive studies
c. Experiments
d. Quasi-experiments
e. *Ex post facto* studies
f. Singe case studies
g. Action research

III Empirical qualitative studies:

a. Ethnography
b. Case studies
c. Instrumental studies

In similar fashion one could classify studies in various broad research fields (and would possibly come up with a similar classification system).

Another way to classify articles is based on their production costs. Anything you produce could potentially be published; the question is often: how many copies should be distributed? SMITH (1990) suggests that there are basically three groups of articles: (1) those with limited results, which are nevertheless surprising and might spark new research (*publish*), (2) papers which mostly repeat work by others (*these should not be published*), and (3) papers which present good ideas, but are badly expressed (*these should not be accepted, but the authors should be encouraged to rewrite and thus make them more comprehensible*).

Based on how often an article is cited, we can roughly distinguish the following five categories of published articles:

Born-dead papers
: These are papers that are almost never cited. Their citation half-life is infinitely short, which means that the topics discussed do not have an audience, the paper is 'indigestible' or it's simply dull. Some papers are simply *bad science chasing a bad idea*, but it takes time until the research community forms an opinion on this. Some might argue that such papers are simply a waste of resources, but this is not completely true. A small proportion of born-dead papers have great potential (*'sleeping beauties'*), but they are too avant-garde, too introspective or too hypothetical. So, it is worth having a record of some good ideas even if we do not exactly know what they mean and how they can be used to solve real-life problems. The real problem is that about two-thirds of all papers belong to this *'born-dead'* category (GARFIELD, 1979; LATOUR, 1987).

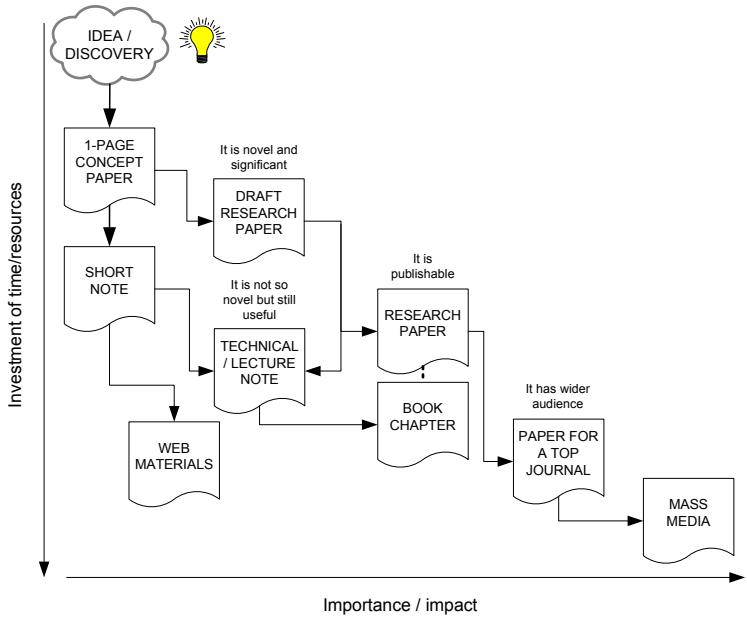


Fig. 1.10 Types of publications in relation to investment of time/resources and potential impact. Ideas evolve or devolve based on the feedback we get from the audience.

Proving-the-known papers

Many papers are reasonably well-written, based on excellent data and the whole project seems to be well organized and conducted, but they are simply not significantly novel (others have already done the work). They can often still be useful, because they retell known theory in a more *'digestible'* way or they are more effective in providing a bigger picture. A paper needs to successfully transfer useful knowledge. So, if an author thinks that he/she can do a much better job then the original author, such a paper will be welcomed. Of course, the author needs to search for and acknowledge the original work (dig into the literature), even if he/she is not initially aware of it.

Promising papers

These are the papers that reveal new discoveries/ideas that are significant for both experimental and applied science. Sometimes, even a badly written paper can be promising. Unlike the *born-dead* and *proving-what-you-know-papers*, authors of promising papers show both talent and dedication to science. Also consider the fact that research groups usually need at least a few years to absorb the ideas laid out in a paper, so such promising papers can eventually get promoted to a higher class.

Most-cited papers
: The most cited (GARFIELD, 1990) or the most downloaded papers are those that have not only proved to be promising, but also those that the research community has shown most interest in. Such papers usually distinguish leading scientists[33] from *run-of-the-mill* ones. In many cases, the most frequently downloaded papers do not need to be of exceptionally high quality (they do not even need to have immediate practical implications), but they should nevertheless tackle the right topics with the right arguments at the right time, and hence provide inspiration for other scientists.

Breakthrough papers
: These are absolute outliers and usually lead to a partial or complete change in an important theory (a paradigm shift). The most famous examples are Einstein's four articles, which he published in 1905 in *Annalen der Physik*, Watson and Crick's *"A Structure for Deoxyribose Nucleic Acid"*[34] and such like. The chances that you will write something like this are extremely low, both in space and time. But you never know.

1.8 Types of journals

Like scientists, journals have the ambition to lead in their field and their performance can also be measured in terms of citations. There are (at least) four types of journals:

The hottest journals
: These are the ones that everyone is dreaming of getting their name printed in. It's hard to define the hottest journals or set a boundary between top and standard journals, but one can certainly make a list of the journals with the highest impact[35]. Hot journals typically have a high rejection rate (>75%), and articles published in such journals are commonly also referred to in the popular media. According to sciencewatch.com, only Nature[36], Science[37] and Proceedings of the National Academy of Sciences[38] (PNAS) are true all-around hot players. Articles in various fields published in these journals are, on average, cited over 50 times per article.

Indexed journals
: Thomson Reuters Scientific monitors journals and, based on some minimum

[33] See for example `http://isihighlycited.com`.

[34] This is one of the most cited research article of all time.

[35] See `http://sciencewatch.com` for an updated list of the hottest journals/papers.

[36] `http://nature.com`

[37] `http://sciencemag.org`

[38] `http://www.pnas.org`

quality criteria, selects journals that it will index. There are three major categories: Arts & Humanities Citation Index, Science Citation Index and Social Sciences Citation Index. For the Natural Sciences, the most important database is the Science Citation Index (SCI), which lists about 9000 journals.

Other international journals
Many journals are not indexed by Thomson Reuters Scientific but still offer a chance to communicate your ideas to a wider audience. In this case it is sufficient if the paper can be easily located and downloaded from Internet. The best-known publication portals are Elsevier's Science Direct[39], Springerlink[40], Wiley[41], Cambridge[42] and Oxford University Press[43]. Note that many journals that provide electronic versions of articles are not included in the Web of Knowledge database and vice versa. So make sure you check out your journal before sending any material.

Local journals
Papers in what we call *'local'* journals are either not accessible to a wider audience or the review process is *'too soft'*. Many journals, even if the review process is rigorous, will remain local because the papers are not written in English. Yes, in science too, globalization (read Americanization) is pretty far advanced. Web of Knowledge is the gold standard for bibliometrics and scientometrics.

Obviously, we would all like to send our papers exclusively to journals that are indexed. On the other hand, sending relatively good papers to journals that are not indexed can be a good investment in such a journal. Remember, it is not Thomson Reuters Scientific that decides which journals are the most important ones — you do! A lot of journals that are now in Thomson Reuters Scientific's database had to go through a rigorous evaluation before they appeared there.

Another way to classify journals is too look at their publishing and archiving policy. For example, the non-profit organization SHERPA (Securing a Hybrid Environment for Research Preservation and Access) distinguishes four categories (see Fig. 5.2 on page 117):

- *Green* — You can archive a pre-print and post-print or provide a publisher's version/PDF
- *Blue* — You can archive a post-print (i.e. final draft post-refereeing) or provide a publisher's version/PDF
- *Yellow* — You can archive a pre-print (i.e. pre-refereeing)
- *White* — Archiving is not formally supported.

[39] `http://sciencedirect.com`

[40] `http://springerlink.com`

[41] `http://onlinelibrary.wiley.com`

[42] `http://journals.cambridge.org`

[43] `http://oxfordjournals.org`

The color basically indicates the accessibility of your work to people who cannot afford journal subscription. More about this topic in chapter 5 (page 113).

Chapter 2
Scientific publishing: myths, ideals and realities

"I think it is important to distinguish fraud — a definite intent to deceive — from bad scientific practice, often a result of inexperience or the current pressure to publish… I think fraud can only possibly be a tiny problem in soil science, whereas bad scientific practice is a much bigger one, but by far the biggest problem we have is a lack of new ideas."[1]

In this chapter we address some major myths and realities related to the way the modern system of science is organized. We take a broad view and suggest some strategies for improving the system. At the end of the chapter, we address the issue of Open Access publishing and archiving.

As Alex McBratney puts it, although fraud and cheating in science represents a serious problem, a much bigger problem for science is *'the bad scientific practice'* — the gray side of scientific work. In our opinion, there are four major reasons for this:

- Ludicrous pressure to publish
- A lack of evaluation of reviewers (and appreciation for their work)
- Fashionable pliability
- Controlled or closed access to the review process and the final products of research.

We all know that problems such as hyperproduction, gift co-authors, self-publication, plagiarizing and poor reviews will continue to exist. The question is whether such practices can be reduced or even prevented? Here are some ideas.

[1] Alex McBratney, joint editor-in-chief of Geoderma, speaking about fraud in Alfred Hartemink's book Publishing in soil science.

2.1 Problem areas

> *"Progress in research is actually highly nonlinear. Papers are often completed as a result of the pressure of deadlines, or the need to turn to other work, rather than ending in a dramatic flourish… the completion of a research paper is therefore often accompanied by negative feelings that after all, not much has been achieved."*[2]

The international Committee on Publication Ethics (COPE) publishes flowcharts[3], which can help editors to systematically manage suspicious publications — those that are the result of fraud and/or copying — and then come up with remedies. The five main problem areas identified by COPE are:

- Redundant or duplicate publications
- Plagiarism
- Fabricated or false data
- Authorship problems
- Conflicts of interest.

Redundant or duplicate publications are possibly the biggest problem in science today. In most national academic systems, scientists are still evaluated by their output (instead of by their impact). People are assessed mainly by the number of papers they publish, so the pressure to publish is rising every day. This has a number of negative effects. We will mention just the three most significant ones:

- *Hyperproduction* — Many researchers find a fruitful topic that is catchy and gets published easily and then go on to publish (very) similar papers in several different journals. This is known as the *hyperproduction effect* (NEWMAN, 2000). Writing more papers on the same topic might be a good idea if it makes a topic better known to different research groups, but if the papers are extremely similar and, especially if the same data and results are emphasized, this cannot be good for science. In extreme cases, hyperproductive authors only change the title of a paper plus a few lines and then publish it in two or more journals.
- *Lobbying and self-publishing* — Many editors, members of editorial boards and even reviewers are biased towards papers with which they have some personal connection. This creates a clear conflict of interest. The extreme case is self-publishing: almost all journals allow the submission of papers of which members of the editorial boards are co-authors. This is bad, both for the editors and for the journals. If an editor publishes his/her paper in a journal while sitting on the editorial board, this will not necessarily be a weak paper. But there is an obvious conflict of interest.
- *Gift authors* — COPE recognizes two main categories of authorship problems (WAGER, 2007): (1) **gift authors** — listed authors who do not qualify

[2] CREEDY (2008) in *"Research without tears."*

[3] `http://publicationethics.org/flowcharts`

for authorship (they may not even be aware of the article and its findings), and (2) **ghost authors** — people who have contributed to the production of an article, but have been omitted from the list for various reasons (e.g. conflicts, hierarchy, fraud). Gift authors are sometimes included to make the list more impressive or to reward them for reasons unconnected with the paper. Mutual favors in this regard are sometimes referred to as *"Mutual resume enhancement."* Often, a person listed as a co-author does not actually know much about the paper and would not be able to defend its content or reproduce it from scratch. Clearly, gift co-authors are only listed because of the benefits of getting published. Consequently, the higher the JIF, the higher the chance that an article contains gift co-authors.

These are many gray areas of science in which it is not easy to categorize specific cases. The borderline between duplication and marginally novel publication can be very *fuzzy*. For example, if the author uses the same data set and the same tools to analyze them, but then reveals a completely new aspect or discovery, then this is certainly acceptable. Most problematic are articles with a somewhat different title and text in the abstract, but results and conclusions that overlap $>50\%$. Papers that present almost the same results and conclusions are known as *"duplicate papers"*[4].

One solution to the problem of gift authors is to limit the number of authors on a paper. For example, the Nobel Prize can be awarded to at most three researchers[5]. In the case of research articles, this would be too few, because it's often good to have more people collaborating on papers.

A simple solution to lobbying and self publishing would be not to allow editors to handle papers where there could be a conflict of interest[6]. However, this is not as trivial as it may seem, because researchers mostly work as editors on a voluntary basis without any financial reward, which means that not many people would edit journals if they were prevented from processing papers in which they have some interest.

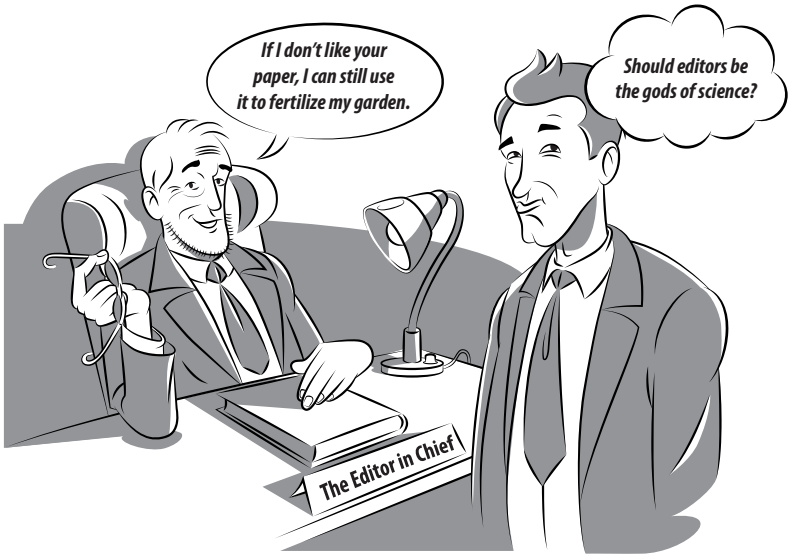
Phoney or gift authors is a problem that has negative side-effects, although at first sight it may not seem to be all that serious. Gift co-authors are basically parasites of science who lack moral values. One may argue that, as long as the first author is authentic, all the others can be phoney, but this situation is much more dangerous than it appears. Firstly, if an author supports a parasite, this means that the parasite will stay in the system of science. After a few years, the hard-working authors will want to apply independently for research funds and then they will have to compete with the parasites, who (on paper) may have similar references. A second more serious effect is that an author who permits gift authors demonstrates a willingness to trade with scientific discoveries, which is obviously immoral.

[4] This can probably be tested statistically, as with patents that are often cross-checked to avoid copying.

[5] See `http://nobelprize.org` under the section *"Statutes"*.

[6] Papers in which the editors are listed as co-authors or papers from departments/units where the editors are employed.

Reputations in the system of science are extremely important. Once it is known that an author is ready to trade moral responsibility for material benefits, then all confidence in this individual will be lost and others will try to avoid collaborating with him/her. The worst-case scenario is when an author accepts the system as such and then one day waits for his/her turn to be a parasite on other colleagues.



Editors usually have the last word in an editorial system. Often, they do not need to justify their actions or decisions to anybody. One of the biggest paradoxes is that the first rule of science rejects the very idea of authority, yet journal editors get exclusive rights to shape science.

The American Physical Society has published a number of ethical guidelines[7] that define a culture — the *'code of honor'* of scientific publishing. The most important points from this code, according to ASCHERON and KICKUTH (2004, p.129-32) are:

- ✓ *Publish substantial and new results only. Avoid re-publication.*
- ✓ *Do not falsify or invent data.*
- ✓ *Avoid plagiarism, respect copyright.*
- ✓ *Limit the list of authors to people who were actually involved.*
- ✓ *Share responsibility/merit with the co-authors (make everybody read and write the paper).*

If you suspect problems with scientific integrity such as fraud, plagiarism, infringement of copyright, incomplete information or improper pressure from

[7] `http://www.aps.org/policy/statements/`

superiors or contract partners, you should first try to contact people within your institute — a department head and/or professional member of staff such as a research advisor — for advice. If this doesn't work, you might consider submitting an objection via the publisher's website (look under *"Report This Content"*) and/or by contacting an international organization such as COPE.

Is someone who just signs bills an author? They shouldn't be. The author of an article must be someone who contributed to the intellectual content of a manuscript by participating in the design of an experiment, in data processing, or in writing and/or editing the article.

The author of an article should only be someone who participated (physically and/or mentally) either in:

1. field/laboratory data collection, and/or
2. data processing and statistical analysis, and/or
3. writing and editing the paper.

If someone is listed as a co-author, this means that he/she made a significant contribution to the intellectual content of the manuscript. A research investigation is not routine work. Hence, a laboratory technician should not become a co-author of the article simply for carrying out routine laboratory analysis that he/she conducts on a regular basis. A co-author of an article should have invested own creativity and original ideas/data. Our experience is that the principal author is usually responsible for producing about one to two thirds of the paper, while the co-authors mostly get involved in the final phases of production. So, if your supervisor, head of project or other superior asks for their name to be listed on the paper, there is still some time for them to get involved. However, if they ask for their name to be automatically listed on the paper without any serious involvement, this is wrong and immoral.

The artificial pressure to publish, with all of its negative side-effects, can be simply avoided by introducing more sophisticated evaluation criteria. Quality is much more important than quantity[8]. Having your name on 20 *born-dead* SCI papers cannot be more important than publishing a single high-impact paper. In fact, in many countries, scientific evaluation teams do not even distinguish between the first and last author.

2.2 Missing reviewers

Reviewers (referees) are crucial for the quality of scientific information. Because reviewers are typically not rewarded or even mentioned for their work,

[8] However, Google derives most of its profit from the brilliant idea of listing sites by the number of times they have been accessed. Quantity sometimes determines quality.

they often deliver a slow or poor service. In fact, reviewers are asked to do high-quality consultancy without any compensation. You could argue that they do benefit by getting early access to papers from colleagues (plus the discipline of reviewing makes them read such papers carefully,which they don't always do otherwise), but the fact is that their work is not directly rewarded.

In an optimal situation, a reviewer will take one day to read the paper and then a few days to cross-check its findings. Such an investment of time is clearly a luxury that few can afford, so reviews are usually done in a few hours. In fact, because the volume of papers published is continually increasing, there is less and less time for proper reviews. Many good reviewers refuse review assignments because they are overloaded with writing papers, or they agree to do reviews which are incomplete or superficial (NEWMARK, 2003; MOORE, 2005). Also, reviewers sometimes grade and comment only on the form and style of a paper and not on its intellectual content.

Fiona Godlee, editor in chief of the BMJ Group thinks that *"scientists are under a lot of pressure on a whole host of things — getting funding and the bureaucracy surrounding scientific research — and peer review is just one other thing, so the more we can do to make it something that they can gain proper recognition for, the better."* (O'DOWD, 2011).

A major problem for many scientific journals is the review process, which is often slow, inefficient, inconsistent, unrepresentative and biased. This is because reviewers are not rewarded for their work or evaluated on their performance. One solution to this problem would be to completely open up the review process and publish both the signed results of reviews and the authors' responses, together with all versions of a paper (*"Open Peer Review"*), so that both authors and reviewers can be properly acknowledged for their work.

Most journals allow 1–6 months for the return of reviews, which does not mean that the reviewers spend that much time reading and thinking about the papers. Unfortunately, when a paper lands on a reviewer's desk, it will first gather dust for some time. Eventually, the reviewer will find time to read it and make comments (usually taking half a day). Often the editor needs to find a replacement for the reviewer because he/she does not respond. Because reviewers are not rewarded for their work, they may find it difficult to justify this investment in time to their employers. Many companies, even government organizations, do not like the fact that their staff spend paid time on reviewing papers, for which only the publisher receives financial benefit.

In the worst-case scenario, reviewers sit on a paper or give it a bad review because it is in their interest. The editor or publisher typically cannot complain because they are not paying for this work. As a result, reviewers often do not feel responsible for their output.

In most cases, an editor will be happy with just two completed review forms. It can easily be shown that a decision based on only two or three reviews can

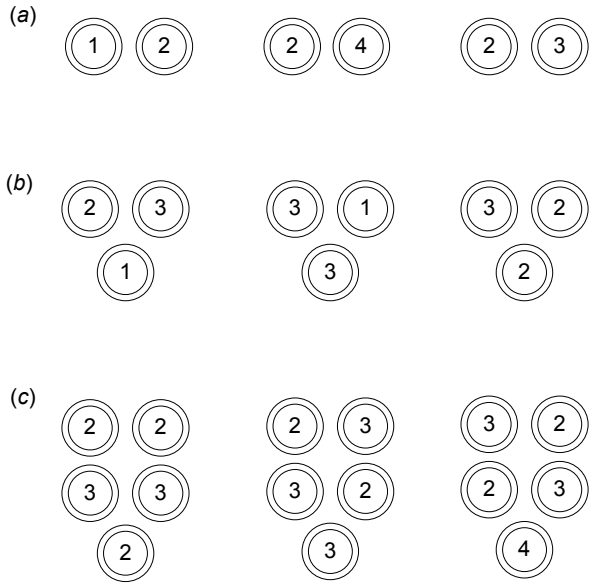


Fig. 2.1 Three simulations of reviewers' decision with different samples. Deciding on an average from a sample of two can lead to some strange outcomes (a). Three is slightly better, but still difficult (b). With five samples (c) we can be probably be more confident. "1" — accepted with minor revision; "2" — accepted with moderate revision; "3" — major revisions needed and "4" — rejected.

lead to fairly poor estimates. In fact, two reviews often contradict each other. Consider a Monte-Carlo simulations of review decisions with different sample sizes. In this example there are four grades for papers: ("1") accepted with minor revision, ("2") accepted with moderate revision, ("3") major revisions needed and ("4") rejected. Now imagine that the grades are based on negative points (0–100), where papers with < 10 negative points are classified as "1", < 40 as "2", < 70 as "3", while the papers with ≥ 70 points are rejected.

Fig. 2.1 shows the results of the Monte-Carlo simulations using x=40, s_x=20 and 2, 3 and 5 for sample size. Note that in the case of only two samples, the decisions can vary from accepted with moderate changes to rejection (Fig. 2.1a). We estimated that, if a sample of two is used, in about 30-50% of cases the decision will differ from the expected one. The situation is a little better for the sample of three (only 30–40%) and much better if a sample of five is used ($< 20\%$). The results of this simulation exercise, of course, depend on how variable the opinion of the reviewers is, but we hope we have illustrated the problem.

A number of studies have been published over the years showing that there is often little agreement among independent referees about whether or not a paper should be published (SCHULTZ, 2010). The major issues connected with why

some decisions are positive and some are negative are: confidentiality, conflicts of interest, editorial freedom and integrity, the lack of uniformity of the review process, etc.

The current problems with poor or delayed reviews could be avoided by rewarding reviewers for their work. This need not necessarily be a financial reward. It would be enough for journals to list reviewers and the amount of work they have done. In fact, editors could monitor how satisfied authors are with the work of reviewers and then, based on certain criteria, promote successful reviewers to become senior reviewers (or at least give them some kind of diploma or symbolic reward). Reviewers could then add such information to their resumés and use it to get greater acknowledgment in their research community. Another (cheap) solution to the problem of biased and unrepresentative reviews would be to ask all members of a society to participate in the review of all papers. This could be organized through on-line editorial systems, in which all reviewers can (at any moment) see the results of the reviews and jointly grade the intellectual and technical quality of an article.

On Scholarpedia: Scholarpedia is the peer-reviewed open-access journal for *living review*-type articles: *"The goal of Scholarpedia is to identify and convince today's Einsteins and Freuds to write encyclopedia articles on their fundamental discoveries so that 100 years from now the best experts will be willing to maintain and update the articles through the process of curatorship."* Scholarpedia articles can be cited and are constantly updated by their curators with contributions by registered members. The job of a curator is to moderate revisions of an article (accept/reject). To publish an article on Scholarpedia you need to get two existing curators to write to the editor-in-chief, and then keep maintaining it.

Scientific journals, in general, could learn a great deal from WIKIPEDIA — the Open Encyclopedia — and the journal version of WIKIPEDIA — Scholarpedia[9] (see box). In WIKIPEDIA, every registered member can at any time edit a topic and see the history of edits and search edits[10]. This saves reviewers a lot of time because they don't need to repeat work, but it also saves the authors time, because they get feedback much faster.

Another way to improve the review process is to reward the reviewers with discounts and/or to assign them some kind of priority status (NEWMARK, 2003). For example, journals published by Berkeley Electronic Press[11] have introduced an innovative concept in which they do not charge submitting authors if they contract to provide a timely review of an agreed number of articles. The authors are charged if they fail to deliver their reviews on time, which seems to be a pretty fair system. Likewise, Scholarpedia ranks authors who contribute to articles so that, once the curatorship of an article becomes vacant, it is automatically offered to the reviewer with the highest index for that article.

[9] `http://scholarpedia.org`

[10] `http://en.wikipedia.org/wiki/Wikipedia:Five_pillars`.

[11] `http://www.bepress.com`

2.3 Fashionable pliability

"What makes things hard to understand is how complicated they are, not how big they are."[12]

The final serious problem with the current system of science is the ease with which authors follow fashionable topics or styles. On the one hand, it is positive to learn from top researchers. On the other, those who imitate other authors forget that in science we need to be cautious and critical about everything. Some authors see their supervisors as Gods of science and blandly repeat whatever they say or write. Their identity is thus lost and they become eternal second authors. Other authors think that, if they choose a sufficiently *'sexy'* topic, this will guarantee them success in getting their papers published (which is often unfortunately true).

To prove that there is a lot of gibberish being published in science today, three MIT students submitted an abstract entitled *"Rooter: a methodology for the typical unification of access points and redundancy"* to the World Multi-Conference on Systemics, Cybernetics and Informatics in 2005. The abstract was accepted for oral presentation and printed and nobody would have complained if the authors had not admitted that they produced this abstract using a computer program[13] that randomly incorporated computer science jargon and produced a grammatically correct yet nonsensical paper (see New Scientist magazine, issue 2496).

A more extreme example is that of Alan Sokal who, in the late 1990s, managed to publish a totally nonsensical paper in a respected journal. Alan Sokal (a Professor of Physics at NYU), succeeded in getting a text advancing critiques of science and rationality common in certain academic disciplines published in a cultural studies journal. Sokal wrote a parody of post-modern science criticism called *"Transgressing the boundaries: towards a transformative hermeneutics of quantum gravity?"*. This was submitted to the cultural studies journal Social Text, without telling the editors that it was a parody. They published it as a serious scholarly article and, when the author revealed the hoax three weeks later, people were very angry with him. Like the genre it was meant to satirize, the article was a mixture of truths, half-truths, falsehoods, non sequiturs and syntactically correct, high-flown language that had no meaning whatsoever. It also contained appeals to authority rather than logic, rhetoric that sounds good but whose meaning was ambiguous, and confusion between the technical and everyday senses of English words (for example, linear, non-linear, local, global, multidimensional, relative, frame of reference, field, anomaly, chaos, catastrophe, logic, irrational, imaginary, complex, real, equality and choice). The article was riddled with significant stupidities and falsities concerning science, which the editors missed.

---

[12] Martin Rees in *"Our Final Century"*.

[13] At `http://pdos.csail.mit.edu/scigen/` you can generate a nonsense paper yourself.

By doing this, Sokal was trying to suggest that in those milieus, some people had virtually no knowledge of the science they were so blithely critiquing. You can read more on the topic in SOKAL and BRICMONT (1997) and SOKAL and BRICMONT (1998). Below is a similar example taken from an original passage (pp. 44-45 and a footnote on pp. 327-328) in *A Short Course in Intellectual Self-Defense* by Normand Baillargeon (BAILLARGEON and SCHMIDT, 2008):

> **The Fox Hypothesis**: Those who pulled off the hoax, which brings to mind Sokal's at the end of the 1990s, formulated what they called the Fox hypothesis, according to which an unintelligible speech, if given by a seemingly legitimate source, will tend to be accepted as intelligible. A corollary of this idea is that using vocabulary that creates the illusion of profundity and erudition can contribute to increasing the credibility of a message. At the beginning of the 1970s, Dr. Fox gave a talk on three different occasions, entitled *"Mathematical Theory of Games and its Application in the Training of Doctors"*. He spoke in front of a total of fifty-five people, all highly educated: social workers, educators, administrators, psychologists, and psychiatrists. His exposition lasted an hour and was followed by a half-hour-long discussion. Then a questionnaire was distributed to the audience to find out what those present thought of the doctor's presentation. All the participants found it clear and stimulating; none of them noticed that the talk was total nonsense, which it was. Dr. Fox was actually an actor. He looked very distinguished and spoke authoritatively and with conviction. But the text he spoke, which he had learned by heart and which had to do with a topic he knew absolutely nothing about, was laden with vague words, contradictions, bogus references to concepts that had nothing to do with the topic, empty concepts, and so on. In short, it was nothing but hot air, contradictions, and pompous meaninglessness.

Young academics and, especially non-native speakers, assume that the best way to get published is to imitate the heavy, unreadable articles they see in many journals and textbooks. Such beliefs are based on a big misunderstanding. What actually gets articles published in top journals and more importantly read, is clear, well-structured, well-argued writing. One underlying issue is that we learn to write in an artificial context, showing teachers and professors what we — and they — already know. We therefore tend not to formulate clear arguments when addressing real-world readers, where the goal is to persuade an audience of the validity of a new solution to a serious problem.

Another common misconception among researchers is that their ultimate career move would be to publish in Nature, Science or a similar journal with a high impact factor. This is a rather naive conception, which is nicely demonstrated by SEGLEN (1997). The correlation between a journal's impact and the actual citation rate of articles from individual scientists or research groups is often poor (Fig. 1.6). In fact, publication in a high-impact journal will not necessarily increase the impact of an article (SEGLEN, 1997). Therefore, you should focus on writing high-quality articles and not on trying to get into top journals at any cost.

2.4 Publishing companies and models

"Who are the most ruthless capitalists in the Western world? Whose monopolistic practices makes WalMart look like a corner shop and Rupert Murdoch look like a socialist? You won't guess the answer in a month of Sundays. While there are plenty of candidates, my vote goes not to the banks, the oil companies or the health insurers, but — wait for it — to academic publishers."[14]

In his review of the current copyright policies in the world, the author and director of the documentary Remix Manifesto, Brett Gaylor, warns us of the increasing risk of corporations taking over control of human culture. For the large corporations ideas are intellectual property that, like any other product, can be commercially exploited to make a profit. Large media corporations are becoming increasingly powerful. In the USA, companies such as Disney, Time-Warner, Viacom, NewsCorp, BMG, and General Electric now own >90% of media holdings. Such large corporations use political lobbies — and their financial strength — to impose more and more control over ideas. *"Patents, innovations and corporate secrets are now guarded like gold."* Companies today try to copyright everything — even broad ideas and legacy information that used to be available in the public domain.

If you extrapolate these trends, it is possible that in some 20–30 years large companies might come to own and control most of human knowledge and thus attempt to completely lock up human culture[15]. Gaylor emphasizes that the public domain needs to be protected and that the control that commercial companies currently have needs to be limited to ensure the continued free exchange of ideas. He builds his argument on the following four premises:

- Culture builds on the past
- The past always try to control the future
- Our future is becoming less free
- To build free societies you must limit the control of the past.

Or as Lawrence Lessig puts it in *"Free Culture"*: *"Overregulation gives dinosaurs a veto over the future; it wastes the extraordinary opportunity for democratic creativity that digital technology enables."*

Something similar can be said for scientific publishing. The best-known traditional **Science, Technical and Medical** (STM) publishers are companies such as Elsevier, Wiley, Springer, and Taylor & Francis, which have a long tradition and often an international network of employees. These companies hold >50% of the global market of STM publishing, which is estimated as between 7 and 11 billion US$ (INFORMATION AND COMMUNICATION UNIT, 2006). For more than 200 years the STM publishing companies have been sharing access to scientific information and have made profits via the system of copyright

[14] Article by George Monbiot. Published in the Guardian 30th August 2011

[15] As Lessig warns us in his book *"Free Culture"*: *"copies in our brain are not — YET — regulated by copyright law."*

ownership (COLLINS, 2005). The market leader, Reed Elsevier, makes nearly 40% of all its profits from the journal business and this profit is significant[16].

The standard publishing model used by the traditional STM publishers is described in Table 2.1. It can be summarized as follows: companies establish copyright over scientific articles in order to ensure financial benefits and they keep the review processes closed to protect their role. Science is then restricted to those who can afford it. More than two thirds of the world's population live in countries that are recipients of Official Development Assistance[17] (ODA). Most people in these countries have no access to *state-of-the-art* scientific literature [18], which means that the gap between the rich and the poor is likely to continue growing. It is not that the commercial STM publishers are responsible for this inequality — they just perpetuate it. If you don't think that this gap is as serious, just try doing science for few months in a country where you can't access most of the recent literature in your field.

Publishing reached a totally new level with the launch of the Internet some 20 years ago. *"Before the Internet, it was inconvenient to retrieve an article that was not available in a local library system. Now it is inconvenient to go to a library"* (COLLINS, 2005). Almost all STM publishing companies have switched to *digital publishing*, which means that most of their products are available on-line as digital media (PDF) and are stored in a database. Big commercial companies, of course, see digital publishing primarily as a new source of profit. But, thanks to the revolution of Internet and a great deal of enthusiasm, some new models are starting to take off: Open Access publishing and Open Access archiving[19].

Table 2.1 shows the characteristics of what we call the *"Open publishing model"* — a model in which articles are immediately published in their original form and made freely available to anyone. In this model reviewers take responsibility for their work, and review — i.e. the pre-publication history — is transparent. Authors keep their copyright and the publisher is closely linked with the research societies connected with the journal.

Is Open Publishing an utopia? Commercial companies have criticized Open Access publishing for not being realistic about their OA author costs. REGAZZI (2004), the managing director at Elsevier, thinks that the author fees of BioMed Central and/or PLoS are insufficient to cover the actual publishing costs. He believes these costs are likely to be the same as those of any traditional publisher, but the final price is lower because the costs are either subsidized and/or the OA publishers underestimate the true costs of sustainable publishing. He thinks OA journals are no threat to commercial publishers, because (1) they are insignificant and (2) they do not really innovate in publishing. *"Open Access or*

[16] The financial reports are available at `http://reports.reedelsevier.com`.

[17] `http://www.oecd.org/dac/stats/daclist`

[18] Thanks to initiatives such as `ahttp://www.research4life.org`, developing countries also have a chance to access research publications at low cost.

[19] To find out more about the Open Access movement, see e.g. Peter Suber's website at `http://www.earlham.edu/~peters/fos/`, and/or read the Berlin declaration on Open Access.

Table 2.1 A comparison between the traditional and the Open Publishing models.

	Traditional STM publishers	Open Publishing model
Copyright held by	company	author(s) of the article
Publication costs (Open Access)	typically in the range 1500–3000 US$	from completely sponsored to 1500 US$
Specification of costs in the bill	×	✓
Time delay from submission to publication	typically in the range 3–12 months	usually minimized (immediate publication)
Line editing and technical support	typically left to authors to organize	on request
Access to pre-publication history	×	✓
Signed reviews	×	✓
Live support	×	on-line status of editors and reviewers visible
Post-publication editing (corrections / errata)	48–hour period for page-proofs	corrections also possible even after publication
Involvement in the activities of research groups	limited	the publisher also organizes conferences and workshops

author-pays publishing solves neither the problem of easy access to scientific information nor the innovation funding problem — OA publishing will neither decrease publishing costs, nor can it generate enough capital to invest in the future development of increased access" (REGAZZI, 2004).

To explore this issue, we can ask whether it is possible to publish quality OA journal articles and books without paying author fees at all. There are few examples, but there are some. The Journal of Statistical Software (JSS, published by the American Statistical Association), for example, is an OA journal that is indexed in Science Citation Index Expanded and its impact factor is increasing (JIF = 2.3 in 2009). What is really interesting about this journal is that there is no charge for submission and no charge for subscription. It gets even better: for both articles and code snippets JSS publishes the source code along with the paper. How do they do it? First, they have minimum costs because they exclusively use Free and Open Source packages (LaTeX for typesetting and Ruby on

Rails for website Content Management System). Second, they distribute only electronic versions of papers, so there are no printing costs[20].

Open Access journals are no different from traditional subscription-based journals — they undergo the same peer-review and quality control as any other scholarly journal. The main difference is that with OA journals: (1) you typically maintain copyright, which means that you can share your knowledge freely, (2) the whole publishing process is usually more transparent.

This book is published via Lulu.com — one of the biggest self-publishing companies in the world. There were NO submission costs for publishing this book at Lulu and we have kept our copyright and revision rights. We could check the price of a single copy long before sending the book to the publisher. Hence OA, either with low submission fees or through sponsored publishing is possible. The question is: how can we make it the mainstream publishing model of the future?

The best-known primarily OA publishers in the world are currently:

- **Public Library of Science (PLoS)** — PLoS is a nonprofit organization of scientists and physicians that publishes articles in biology, medicine, genetics, computational biology and related fields. The OA costs per article are in the range of 1500-2000 US$, depending on the journal. These costs are lower than those of STM publishers, partly because PLoS has received millions of dollars of charitable support.
- **BioMed Central (BMC)** — BMC (now owned by Springer) currently charges between 500 and 1750 US$ per article.
- **Hindawi Publishing Corporation** — Founded in 1997 and based in Egypt, Hindawi maintains 200+ journals in various fields from Agriculture to Physics and Neuroscience. It has a large number of editors spread around the globe.
- **Copernicus Publications** — Copernicus maintains some 40+ journals, primarily for the Earth sciences and geosciences.
- **Molecular Diversity Preservation International (MDPI) Publishing** — MDPI is an organization for the deposit and exchange of molecular and biomolecular samples located in Basel, Switzerland. Established in 1996, it hosts some 40+ Open Access journals, mainly in the field of molecular biology and bio-sciences.
- *Other specialized OA publishers*[21] — In the USA and Europe there are now several scholarly publishing houses that can completely accommodate OA publishing. For example, the American Meteorological Society and the Institute of Electrical and Electronics Engineers.

[20] Books are also increasingly published as OA materials. A number of *"Free"* books can be found at `http://www.gnu.org/doc/other-free-books.html`.

[21] For a complete list see: `http://www.doaj.org`.

OA is slowly becoming accepted by both commercial and academic publishers. Even commercial publishers have recognized the potential of OA publishing and its ethical logic. The main barrier is still conservatism among authors and publishers alike. OA is also often misunderstand. To clarify the main principles of OA, BioMed Central, one of the leading Open Access publishers has published a lengthly response to the most prevalent anti-open-access arguments — Misleading Open Access Myths[22].

There are also two moral dilemmas in the current toll-access publishing system. First, science should not be a luxury that's restricted to an elite. Second, why should we have to pay for the results of research that is funded by public money? As KLEINER (2011) puts it: *"If the public is funding the research, the public should be able to see it."*

On self-publishing: Self-publishing is publishing outside formal organizations i.e. without professional publishers. The main problem with it is that it often leads to *'gray'* literature. This is the type of literature that lacks strict bibliographic control or professional layout and can therefore disappear in a few years. Self-publishing typically means low-budget production, because individual authors start with limited resources. Another problem is that self-publishing suggests that the authors do not appreciate the opinion of the research community or that the work is unchecked. The whole idea of peer review is that other people read someone's work and then give feedback and evaluate it. However, as we argue elsewhere in this book (page 37), the peer review system is in urgent need of drastic overhaul. Hence most publications in STM journals are also the result of self-publishing in a way — journal editors do not try to improve or help redesign articles, but are primarily interested in their ranking (i.e. minimum potential negative publicity). Most peer reviewed (commercial) scientific publications express only the opinions of authors, are often poorly checked, and come with no warranty.

Self-publishing should not present science with any problems as long as the values are right. REES (2004, p.85) agrees that *"any potentially epochal claim, provided it is openly announced, will be guaranteed to attract wide scrutiny from the international community of experts. So it doesn't matter a great deal if formal review is bypassed, provided that there is no impediment to openness."* In other words: open publishing is probably more important to science than peer review.

Many large STM publishers now provide an OA option by charging authors either a fixed amount or an estimate that is based on average costs[23]. Elsevier, for example, charges about 3000 US$ for open access. A number of institutions[24] have signed a special agreement with Elsevier to meet the OA fees on behalf of their authors. Springer has started a user-friendly service for Open Ac-

[22] `http://www.biomedcentral.org/openaccess`

[23] See `http://www.sherpa.ac.uk/romeo/PaidOA.html`.

[24] `http://www.elsevier.com/fundingbodies`

cess publications called *"Open Choice"*[25]. The idea behind this is to simplify OA publishing: Springer signs agreements with universities and libraries, so that the bill for OA goes to the organization and not to individual authors. This *"green road"* to OA publishing may well represent the best business model to follow until the wider scientific community comes to accept the logic of OA. Springer also hosts a number of Open Access journals[26], and is currently the leading STM company in the field of OA.

THE WELLCOME TRUST (2004) has estimated that the costs for a good-to-high-quality OA article would be about 2000 US$. Their report further recognizes that *"an author-pays system has the potential to be more economically efficient, both in terms of the allocation of resources between competing uses and in the level of total system costs."* It is likely that the provision of open, electronic archives will effectively create an open access system for readers *"which could fatally damage subscriber-pays systems in the long term"* (THE WELLCOME TRUST, 2004). While the pros and cons of OA publishing are still being debated, we believe there are no valid arguments against OA archiving (see next page), especially when you think of the benefits it brings to developing countries. EVANS and REIMER (2009) discovered that, in developing countries, free-access articles are much more likely to be cited and are likely to make more impact than articles in commercial journals.

Internet and its side-effects — Internet is considered to be one of the most promising prospects for humanity (GLENN *et al.*, 2009). The advent of worldwide, decentralized communication epitomized by the internet and cell phones has been a pervasive democratizing force (KURZWEIL, 2005). Think of the role they have played in the popular revolutions in Tunisia, Egypt, Yemen and Syria. Nearly 25% of humanity is already connected to the internet. If one extrapolates current internet growth trends (internet traffic growth doubles every 6 months; bits/dollar efficiency doubles every 12 months; the internet router/switch max speed doubles every 6 months), then it is easy to see that in 5–10 years almost everyone will have efficient, inexpensive access (Fig. 2.2). In such a future, open access to science should not cost a fortune (as the commercial publishers argue right now).

Another good example of an OA publisher that supports open publishing is **Copernicus**[27], a not-for-profit publishing corporation closely linked to the European Geosciences Union (EGU). The model that Copernicus uses is truly innovative. First, submitted papers are immediately published as discussion papers, which means that there is no delay between submission and publication of research results. Second, reviewer's reports, authors' responses and any other comments — which anyone can submit, whether anonymously or with contact details — are also published immediately. Third, Copernicus is closely involved in the organization of conferences and similar events for the EGU. Copernicus

[25] `http://www.springer.com/openchoice`

[26] `http://www.springeropen.com/journals`

[27] `http://publications.copernicus.org`

charges various author fees, depending on the journal, which are fully specified in the bill sent to authors.

The following five practical advantages of OA publication are worth mentioning:

(*a*) *You allow open access to your work*,
(*b*) *Your article is immediately available on-line*,
(*c*) *You own the copyright for your work*,
(*d*) *You can track who is reading your paper and when*,
(*e*) *You can print out as many copies of your article as you need and e-mail copies to anyone you like*.

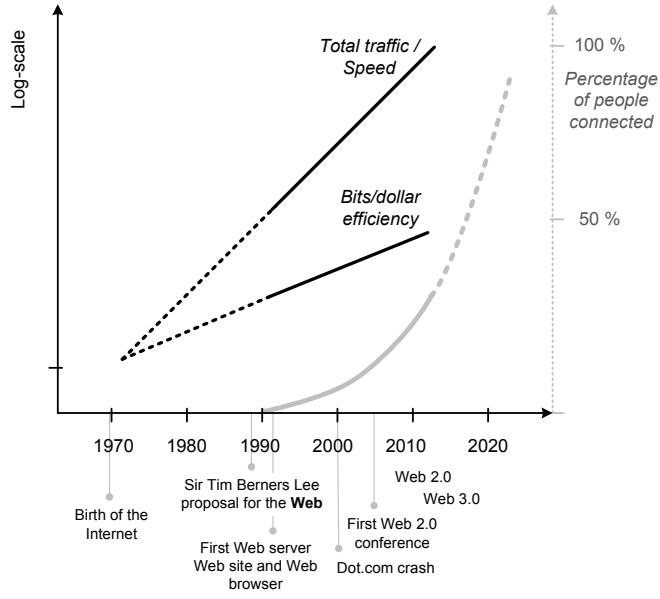


Fig. 2.2 Evolution of the internet follows an exponential growth curve. There is every reason to expect that in the 5–10 years almost everyone will have efficient inexpensive access.

Finally, articles can easily be made accessible to the scientific community through **open access archiving**, by interlinking repositories of publications or by publishing OA digital copies after an agreed period of time. The UK Government Select Committee Report on Science Publishing[28] says: *"We recommend that the Research Councils and other official funding organizations mandate their funded researchers to deposit a copy of their articles in their institution's repository within one month of publication... as a condition of their grant,"*

[28] Recommendations published at `http://epublishingtrust.org`.

and the US government has recommended that research funded by the National Institutes of Health be likewise archived. Archiving already published research in interoperable institutional archives greatly benefits global science at virtually zero cost. This can be done now, without changing established publishing practices (INFORMATION AND COMMUNICATION UNIT, 2006). For developing countries this creates enormous opportunities, especially in agricultural and medical research[29].

OA publishing and archiving models will put pressure on commercial publishers. Users should clearly benefit from this. They should pay less for higher quality scientific information, and this will also benefit peer review, which should become more transparent.

2.5 Publishing efficiency

The fact that big traditional SMT companies now publish a price for OA allows us to track scientific **information production efficiency** (IPE) for any research article. Theoretically speaking, IPE can be expressed as:

$$\text{production efficiency} = \frac{\text{total costs}}{\text{information bits}} \tag{2.1}$$

This parameter is not easy to estimate per article. Both the numerator and the denominator are complex variables, consisting of many types of input. The total cost of producing scientific information consists of at least three groups of costs:

- *The cost of doing the research* — These are the costs of running research, which typically cover the following:
  - Staff salaries
  - Access to publications (library costs, subscription fees)
  - Rooms and laboratories
  - Lab equipment and materials
  - Traveling costs (conferences, workshops)
  - Training and innovation
  - Software and IT (software licences, hardware, network and maintenance)
  - Administration and legal issues
  - Insurance of people and equipment.
- *Publication costs* — These comprise various article processing services including:
  - Administration and communication
  - Peer review

[29] Steven Harnad: *"Self archive unto others as you would have them self archive unto you."* See also `http://www.eprints.org/openaccess/`.

  - Formatting
  - Indexing (DOI, ISBN, etc.)
  - Archiving
  - Web hosting
  - Marketing
  - Innovation / new projects
  - Printing.

- *Post-production costs* — The cost of continuing the research; these include:
  - Promotional materials, brochures, reports and manuals, websites
  - Promotional events (workshops, training courses)
  - New acquisitions (new staff, new project proposals)
  - New systems (research and development).

Total information production costs can be estimated as the sum total of the hours spent on producing the paper multiplied by unit costs (per author), plus publication costs. Estimating total production costs is difficult, because it is not easy to quantify how much time authors have spent on producing an article. Note also that many research groups underestimate the importance of post-production costs, which are often crucial for ensuring their continuity.

Instead of calculating the IPE, we propose here to use a simplified measure — **publication efficiency** (PE), which can be expressed as:

$$\mathrm{PE} = \frac{\text{OA costs}}{\mathrm{CR}} \tag{2.2}$$

Or in other words: PE = publication costs per unit of impact. The lower the PE the better of course.

PE is much easier to estimate and it provides a quick measure of how efficient an information product is. For example, suppose someone publishes an article in an SMT journal, and pays 3000 US$ for OA publication. This paper could, for example, achieve a CR of 20, which would make the PE of 150 US$ per citation. If another author manages to get the same CR in an OA journal that charges only 500 US$ per article (PE = 25 US$ per citation), then this means that the paper published in the 500 US$ per article journal is six times more efficient.

Obviously, PE cannot be derived for articles where the cost of OA publication is 0. To avoid such problems, each journal should publish some realistic number indicating the costs of publishing an article, even when these costs are 100% subsidized.

Ferry Dizadji of the European Commission has evaluated journal subscriptions and citation statistics and has suggested a similar **Relative Cost Index**:

$$\mathrm{RCI} = \frac{\mathrm{CPI}}{\mathrm{CPI}_{OA}} \tag{2.3}$$

where CPI is the Composite Price Index, i.e. the geometric mean of the Price per article and Price per citation, and CPI_{OA} is the CPI of a non-profit journal in the same category. Again, the lower the RCI the more efficient the journal.

Academic institutions should track their research and look at concrete relevant data such as publication efficiency (the cost of publishing divided by scientific impact) and/or the relative cost index, and then use such parameters to encourage researchers to produce influential work at the lowest possible cost. It is very likely that expensive publishers will also score well on PE because good citation statistics often result from high-quality production, but the point we are trying to make is that such information should be made transparent.

A major misconception in science is that the quality of articles is mainly due to the journal's IF, i.e. its name. What we want to emphasize in this chapter is that the quality of research articles is primarily a result of good ideas and good research (see further Fig. 5.3). Production and post-production are important, but the quality comes from the authors. Measures such as JIF suggest that the scientific quality comes from journals i.e. publishers, and that they alone should be used to evaluate the quality of published research. Fig. 1.6 clearly shows that JIF has not much to do with the individual impact of articles and so should not be used to evaluate researchers.

Note that we do not want to imply here that scientific publishers should be replaced with *quick-and-dirty* systems. On the contrary, there would not be any science without companies such as Elsevier, Springer, Taylor & Francis and Wiley. However, our general impression is that access to science could be extended and that the time it takes to get from research to application could be considerably shortened, to the benefit of all.

Part II
Guide for authors

Chapter 3
Producing scientific information

So far, we've discussed some of the problems with the current system of science. In this part of the book we provide practical tips 'n tricks on how to produce high-quality papers. The first thing you need to realize is that it's all about having a good idea and a lot of discipline.

Papers are born from ideas, i.e. when an author intuitively senses an important discovery. In principle, every idea/discovery can potentially lead to publication (as illustrated in Fig. 1.10). But not all drafts are publishable. So you should first ask yourselves whether your work is truly novel and whether it has a large enough potential audience. In many cases, you may need to face the fact that your ideas are simply not good enough for a top journal (or perhaps not even good enough for any journal). Such work can still be disseminated, but in a different format. Once you are sure, however, that you want to produce a research article, you need to be systematic. Producing scientific publications generally takes place in five main stages:

- *Design and preparation of the scientific publication* (experimental design and agreement among co-authors)
- *Actual research work* (data collection, data analysis, scientific writing)
- *Product filtering* i.e. peer-review, line editing and production of graphics
- *Publication and dissemination*
- *Post-production* (mainly product marketing).

In the following sections we describe these stages. We first focus on preparation and data collection. In chapter 4 (page 71) we focus on the techniques and skills needed to write papers, and in chapter 5 (page 113) we provide some pre-submission checklists. We highlight both threats and opportunities and illustrate these with some examples.

3.1 Design and preparation

> *"I believe in intuition and inspiration. Imagination is more important than knowledge. For knowledge is limited, whereas imagination embraces the entire world, stimulating progress, giving birth to evolution. It is, strictly speaking, a real factor in scientific research."*[1]

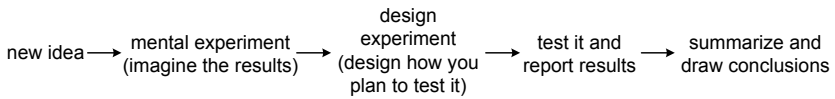
People generate ideas in different ways. One natural way to come up with ideas is:

- *Gather raw material — specific & general — and combine it kaleidoscopically.*
- *Digest this material mentally.*
- *Drop the problem and do something completely different.*
- *Experience the Eureka moment, when the creative idea appears as if out of nowhere.*
- *Expose the idea to criticism i.e. test it against the real world.*

The creation of ideas is usually a non-linear process. It often happens as the result of complex mental processes that are still not fully understood. Nevertheless, we can create an environment that's conducive to generating ideas. Life on earth developed in physical and chemical conditions over 2 billion years ago. Simple molecules just needed enough energy, water, chemicals and radiation; the rest they figured out themselves. In the same way if we expose a human brain to inspiring ideas and provide materials/tools that it can play with, it is very likely that these will lead to new ideas and hence to new creations. For practical tips on how to improve your creativity by improving your working conditions, see section 3.6 (page 66) on coping with stress.

For Prof. McBratney, University of Sydney, the typical cycle of producing a scientific information package is:

1. *Generate an important idea*
2. *Develop it (design the experiment)*
3. *Think how to test it (choose the right statistical method)*
4. *Collect data (measure)*
5. *Test it*
6. *Publish it*
7. *Move on to the next idea…*



Also for CREEDY (2008), the pre-writing stage is crucial to improving efficiency. He offers a number of practical tips:

[1] Albert Einstein in *"Cosmic Religion: With Other Opinions and Aphorisms"* (1931), p. 97.

- *Attach a time schedule to your writing plan* (aim to finish with several weeks to spare before the deadline).
- *Start writing immediately* (write as you go along).
- *Establish good working habits and use your time efficiently: work on a bibliography or on tidying up graphics and supplementary materials when you can't work on the main parts of your paper.*
- *Arrange regular meetings, be flexible and revise your plan when necessary.*

This last tip needs to be emphasized. Science requires flexible agendas; you can't make progress with strict routines. A *live* agenda that can be iteratively adjusted based on initial results and difficulties typically leads to faster progress, because science can be extremely unpredictable (CRUMP, 2002). Problem solving should dictate the way we design science.

It's also important to emphasize that many researchers (especially beginners) put themselves under unnecessary stress by seriously under-estimating the time/effort needed to complete project phases. CREEDY (2008, p.22) thinks that we should be much more pessimistic: *"when planning research projects, produce a generous estimate, fully allowing for the fact that everything takes longer — then double the time and add some more for good measure."*

Wikihow (a manual that anyone can edit) lists the following key tips on *"How to Conduct Academic Research"*[2], i.e. how to write an essay or review paper on a research topic:

(*a*) *Design your research*: Determine your research topic/question. Understand the difference between primary (original) and secondary (review) research. Determine your scope and time line. Write a research question, which reflects a real problem that needs to be solved. Ideally it should contain variables or other relationships that can be tested.
(*b*) *Read the relevant literature*: Learn how to find useful sources. Collect some possible sources and begin reading in detail. Find a method to take notes on what you read. Continue to consider new sources.
(*c*) *Evaluate*: Evaluate the sources you use. Keep your research question in mind. Your source material must help you establish your thesis. Be selective. Don't be tempted to write an exhaustive WIKIPEDIA-style review of the topic.
(*d*) *Formulate the thesis*: Write your tentative **thesis**[3]. Think of how to express your point in a single, complete sentence. Make sure this sentence states your opinion.
(*e*) *Begin writing*: Begin writing your first draft. First sketch a rough outline, which explains the problem you are tackling, a research question (or series of questions) designed to solve that problem, answers to these questions, the implications of those answers, and possible next steps for research.
(*f*) *Revise it*: Continue writing your first draft with quotes (or paraphrases) from relevant sources, and then revise it.

[2] `http://www.wikihow.com/Conduct-Academic-Research`

[3] This is a single statement of your position on the research question.

(*g*) *Finalize it*: Prepare the final draft. Strictly follow the format of the target journal, by checking its handbook or a general stylebook (e.g. TURABIAN (2007)). This includes: title page, page setup and numeration, citations, bibliography style, visuals, sections and titles, etc.

For ROSSITER (2009), reporting on research follows Caesar's proverb *veni*, *vidi*, *vici*, or in other words: I came and applied some methods to attack the problem (*veni*), I saw the following results (*vidi*) and I can now draw some significant conclusions (*vici*). One victory leads to another battle, of course.

For Alex McBratney[4], the keys to success in scientific work are:

- *engage in deep reflection*
- *talk to people*
- *use mind-altering devices*
- *no Ipod!*[5]

Assuming that you have an idea (and frankly, it's not important how you got it as long as it's your own and it's a good one), this needs to be converted into a clear, concise proposal.

3.2 The one-page concept paper

The first tip for producing relevant, credible, readable scientific information is to carefully plan the whole thing right from the start. Pitch your idea as if you were trying to convince a journal editor of the value of your research. We call this a one-page concept paper. As we mentioned earlier, editors take just a few minutes to decide whether a paper that's been submitted should go to peer review or be rejected. This paper should include the topic, the authors and their roles and responsibilities, your main ideas and assumptions, a broad picture of the experimental setup and a time-line with phases and deliverables. Once the main thrust of your paper has been established, it's much easier to organize the production of the paper. Think of it as a small project.

These are some issues that you should definitively consider when preparing a one-page concept paper:

✓ What do you want to *'sell'* with this paper? What is the problem that you're addressing and what is the key research question that will lead you to a solution? The research question (and its answer) is the basis of all credible science. Blaise Pascal: *"One cannot really be considered as having a research topic until it can be expressed in the form of a succinct question."*

[4] Keynote talk at the Pedometrics 2007 conference.

[5] Hopefully Apple will not take this remark as anti-marketing. The fact is that many modern entertainment devices tend to capture a lot of our attention, with the risk that we lose focus.

- ✓ Is the topic really[6] novel?
- ✓ Who is it intended for: a specialist or broad audience?
- ✓ What will be its strong aspects?
- ✓ How are you going to prove your hypothesis and is this proof going to be convincing?
- ✓ Will you be able to organize the experiment and data processing (resources, support)?
- ✓ Who will be first, second author, etc., and what will be their responsibilities?
- ✓ In which form do you want to publish it?

The most important step in starting a paper is to produce a one-page concept paper. This should include: an important issue (the topic), a clear problem statement, the authors, their roles and responsibilities, your main ideas and assumptions, a broad picture of the experimental setup, and a time-line with deliverables.

Although it is a good idea to select a target journal early on, at this stage you should first focus on the quality of your research and not think too much about the impact factor of the journal or the number of publications you can produce from your results. Keep in mind that a good paper is one that makes an impact, i.e. one that will be widely read and used by many people to further their research.

As we explained in the first chapter, research publications typically focus on one (or more) of the following:

I. New discoveries (about us and our environment), which could include (CREEDY, 2008):
   - ○ New empirical regularities
   - ○ A new theory, and/or
   - ○ Improved understanding of or fresh insights into a problem.

II. New technological developments
III. Solving open mysteries
IV. Systematization and synthesis of existing knowledge (overview and/or review).

Try to distinguish in which category your paper falls. Perhaps it's all four. In which case it would be rather complex to write such an article. Maybe your work should be split into several articles?

[6] Often we are sure that the topic we are working on is completely novel. We then find out that it has already been discussed and described, sometimes as long as 50 years ago.

3.3 Review your results and repeat the experiment

"I only trust those statistics that I've falsified myself."[7].

Now you have your master plan, you can proceed with the collection of data, i.e. carry out experiments. The initial results confirm your expectations and you are excited about the whole thing. You would like to publish it as soon as possible. At this stage, it might be wise to review your results and even repeat the experiment several times. Sleep on it. As John Tukey correctly puts it: *"the combination of some data and an aching desire for an answer does not ensure that a reasonable answer can be extracted from a given body of data."*

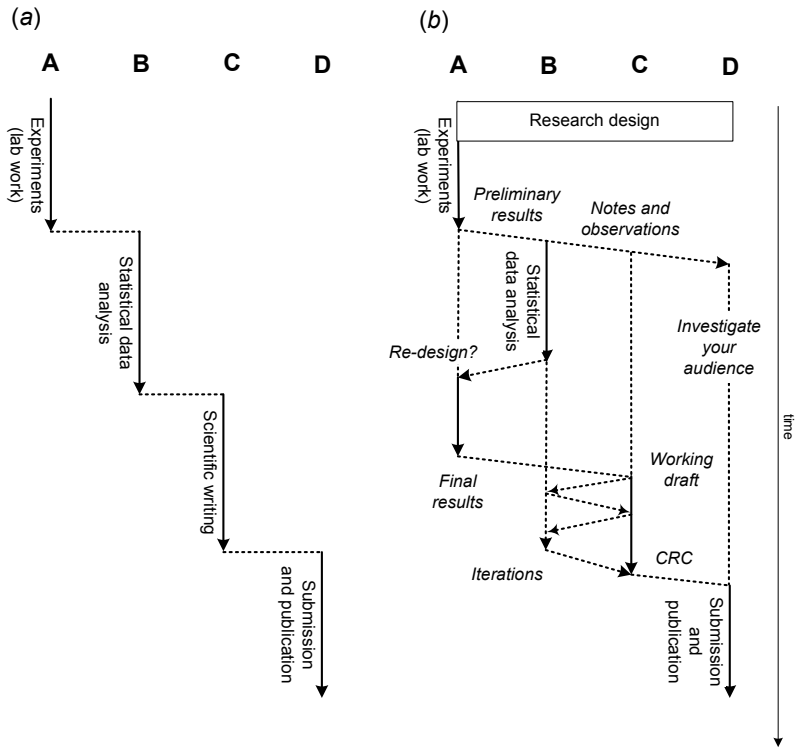


Fig. 3.1 Scientific paper production phases — lab work (A), analysis (B), writing (C) and submission (D): (a) the 'linear' approach (which almost never works), and (b) the iterative, parallel approach. In practice, evolution of scientific information up to Camera Ready Copy (CRC) is, in fact, highly non-linear with many iterations, resets and re-designs.

[7] Winston Churchill

The worst-case scenario is that you get your paper published and then find out that some aspects or elements were incorrect or wrongly interpreted. Once people find out, you get a bad reputation and you will have a much lower chance of publishing similar papers in future. So, although you scored a publication, you've damaged your career. Prof. McBratney suggests: *"spend more time thinking how to test a model before it's too late."*

The most critical parts of your article are likely to be the evidence and interpretation. Spend time thinking how to make your arguments more convincing. Reconsider your results and, if possible, repeat the experiment several times.

This brings us to another big problem of modern science — impatience. Authors are often impatient to publish, so they hide unexpected findings or things that they cannot explain. Sometimes, research projects can lead to what is called *'negative'* results — proving that the proposed methodological or technological improvement does not fit expectations or does not help solve some practical problem. CREEDY (2008) points out that even such disappointing results can be useful and should certainly not be dismissed. In fact, some of the best articles in the history of science focused on something that DID NOT WORK. Are you aware that the most significant discoveries[8] in the world happened unexpectedly, through serendipity or even error? Really important ideas only become clear in retrospect. Great ideas might be emerging right now, but we don't know it. In other words, if you are too sure about the results you expect and if there's too much routine in your work, do not expect to discover something great (see also the rules of science on page 5). For the same reason, always be very flexible and ready to adjust the key topic of your article, depending on what you and your co-authors think is the most significant discovery.

3.4 Investigate your audience

Once you've done several tests and got the same results repeatedly, you can be confident about your discoveries. However, you should not immediately complete the paper. Now is a good moment to investigate your audience, i.e. those who will read and evaluate your work: focus on the audience and readability of the paper. This is nicely emphasized by GOPEN and SWAN (1990): *"An academic paper cannot exist without the interpretation of each reader. If the reader is to understand the writer, the writer has to know what the reader needs. We can't be sure that even a single sentence we write will mean the same to every*

[8] For example, electricity, telephone, Röntgen rays, cosmic background radiation, etc. See a book on this topic by NEWMAN (2000).

reader; all we can do is increase the chances that most readers will interpret our writing the way we intended."

The best way to find out how potential reviewers will receive your paper is to communicate some preliminary results at a research conference or seminar. Communicating your preliminary results and key ideas to potential reviewers will give you insight into what they see as strong points and what they criticize. You can get such feedback in few hours (if you send a paper to a journal, you will have to wait for months). By giving seminars you can also practice putting your thoughts into words and then into arguments.

If you do not get any questions about your work, this is a bad sign. Either your colleagues are not interested in the topic, or they have difficulties understanding it, or you have not emphasized the key points in your presentation. Also, if you offer too many ideas/results (even good ones), this can tire an audience and they will not be receptive to your work. The same will probably happen with the paper. Sometimes, throwing things out of the paper really helps — *less is more*! Many investigations[9] have shown that shorter, more focused papers generally have a higher impact. One of the reasons why short is better is because the authors have had to put more work into compressing their work. As Blaise Pascal once said: *"Je n'ai fait celle-ci plus longue que parce que je n'ai pas eu le loisir de la faire plus courte"* or *"I have made this (letter) longer than usual, only because I have not had the time to make it shorter."*

Research conferences are also a good place to find out more about the topics that your colleagues are working on. It's not only important to find out what others think about your ideas, it's also important to know what other people are working on at the moment. The best scenario is that your topic (research problem) is discussed heatedly by many other scientists, i.e. it's *'hot'*. This is definitively a sign that you should start preparing the first draft of your paper.

3.5 Ten simple rules

> *"First principles, Clarice. Simplicity. Read Marcus Aurelius. Of each particular thing, ask: what is it, in itself, what is its nature...? "* [10]

PLoS has published a collection of *Ten simple rules*[11] for various aspects of scientific work. The ten rules for doing best research according to Hamming are (ERREN *et al.*, 2007):

1. *Forget modesty and say to yourself "I want to do something significant."* (Go Big or Go Home)

[9] See for example the work of HARTEMINK (2002) about publishing in soil science.

[10] The first simplicity principle of Hannibal Lecter that helped agent Clarice Starling solve the case of a serial killer. From *"The Silence of the Lambs"* book by Thomas Harris.

[11] `http://collections.plos.org/ploscompbiol/tensimplerules.php`

2. *Prepare your mind — luck is a marriage between opportunity and preparation.*
3. *Start publishing young.*
4. *Brains are not enough; you also need courage.*
5. *Make the most of your working conditions — don't blame the tools.*
6. *Work hard and effectively.*
7. *Believe and question your hypothesis at the same time.*
8. *Focus on what is important for society.*
9. *Be committed to your problem.*
10. *Leave your (office) door open — don't get too isolated.*

Likewise, BOURNE (2005) compiled ten simple rules for getting published:

1. *Read many papers, and learn from both the good and the bad work of others.*
2. *Learn to be objective (as the journal editors) early. The more objective you can be about your work, the better that work will ultimately become.*
3. *Look at the masthead of the journal in which you plan to publish. Good editors and reviewers will be objective about your work.*
4. *Learn to write well in the English language.*
5. *Learn to live with rejection.*
6. *Do not ignore the essential ingredients of good science/reporting: novelty, comprehensive coverage of the literature, good data, good analysis and thought-provoking discussion, good organization of the document, appropriate use of tables and figures, right length, writing to intended audience.*
7. *Start writing the paper the day you have the idea of what questions to pursue.*
8. *Become a reviewer early in your career.*
9. *Decide early on where to try to publish your paper.*
10. *Quality is EVERYTHING. Better publish one paper in a quality journal than multiple papers in lesser journals.*

Such rules of thumb won't necessarily work for every field of research or for every individual, but they are based on decades of experience from a variety of research fields and cultures, so are certainly worth considering. Although it might seem simplistic to reduce everything to simple rules, these are essentially sound.

So, the first question you need to answer is: how ambitious are you? Obviously, if you want to do top-class work, you will need to work hard: six of the ten rules refer to preparation, commitment, and focus. On the other hand, there's no need to be obsessed about your work. In fact, the best ideas come from a healthy body and soul and not from obsession and isolation. Participation in regular sporting activities or other non-intellectual pursuits provides a valuable diversion from work (CREEDY, 2008). For examples, sports such as jogging, swimming or cycling are excellent activities to get fit while allowing your unconscious mind to continue processing intellectual problems (see section 3.1 on generating ideas).

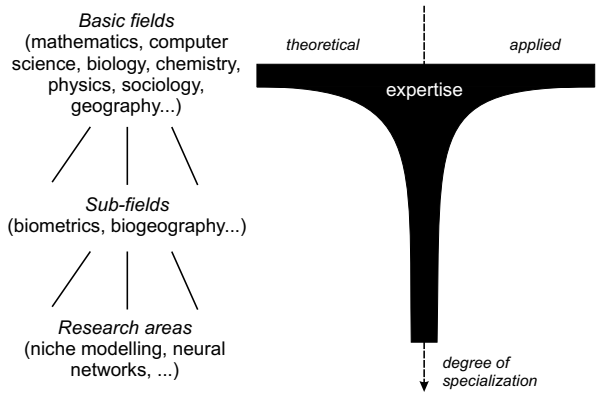


Fig. 3.2 The concept of T-shaped skills: a modern researcher is expected to be a hybrid between a generalist and a specialist, with equal ability in the applied and theoretical sides of knowledge.

Group sports are also useful for developing social skills and strengthening your stamina and crisis management skills. Another useful way to relax intellectually and get inspiration is to read popular science books and science articles in the media (as long as these articles are based on systematic research and not speculations). Seminars can also be very inspiring, even when the topics discussed are very different from your own work.

Many sociologists and managers think that intellectual workers with T-shaped competencies are better at coping with real-life problems (see Fig. 3.2). Thus, investing in your general knowledge could be beneficial for your success as a researcher.

Rik Leemans, editor in chief of the journal Current Opinion in Environmental Sustainability and author of several influential papers, when asked about the secrets of success for writing winning articles, refers to the following *spices*:

✓ *the team*
✓ *the guts to be innovative*
✓ *effort spent on producing a new tool and/or new large database*
✓ *effort spent on promoting it (through workshops and press releases).*

Note that teams increasingly trump *solo* authors: teams typically produce more frequently cited research than individuals do, and this advantage has been increasing over time[12] (WUCHTY *et al.*, 2007).

[12] In the old days solo authors were more likely to produce exceptionally high-impact research.

Furthermore, for Leemans, the key to success is to produce papers that are winners in at least four categories:

- *it's a story*
- *it's interesting*
- *it's clear (easy to grasp)*
- *it will sell*.

Leemans suggest that, in order to improve readability, authors should always use the language of their target audience and define or *'translate'* unusual terms.

The key to producing highly influential articles is: (1) focus on a topic that is relevant, (2) get the best co-authors on your team, (3) demonstrate your points by using clear examples, (4) package your paper with accompanying materials (posters, software, web-sites, video-demos, promotional materials).

The European Association of Science Editors (EASE) has produced a list of guidelines for writing research articles and other scientific publications[13], as a result of long discussions on the EASE Forum and during the EASE 2009 conference in Pisa. Here is a summary:

1. *Do not begin drafting the whole paper until you are sure that your findings are reasonably firm and complete*, so that you can draw sensible and reliable conclusions.
2. *Choose the right journal for your manuscript before you start writing*.
3. *Do not submit articles that are not 100% complete*.
4. *Follow the logical macro-structure suggested by the publisher* — information is interpreted more easily if it is placed where readers expect to find it.
5. *Do not include information that is not relevant to your research question(s)*.
6. *Do not copy and paste* (substantial parts) *from previous publications*.
7. *Do not repeat information in the article* (with the exception of the abstract, figure legends and concluding paragraphs).
8. *Reduce the length wherever possible* (delete obvious statements and other redundant fragments).
9. *Replace long scientific terms and expressions with abbreviations*.
10. *Express your doubts if necessary but avoid excessive hedging*.
11. (Unless required otherwise by the editors), *use numerals for all numeric data*, i.e. also for single-digit whole numbers, *except for zero and one* (if without units), and in other cases where misunderstanding is possible. In numbers exceeding 4 digits to the right or left of the decimal point, use thin spaces (not commas) between groups of 3 digits.

[13] `http://www.ease.org.uk/guidelines/`

12. *Clearly distinguish your original data and ideas from those of other people and from your earlier publications.*
13. *Check that you are using correct scientific terms*. Define every uncommon or ambiguous scientific term at first use. Avoid colloquial and idiomatic expressions. If in doubt, replace unfamiliar terms with easily understood terms with a similar meaning.
14. Add the original names of places to lesser known geographic names.
15. *Write compact, cohesive and logically organized text*. Each paragraph should preferably start with a topic sentence, with the next sentences fully developing the topic.
16. *Do not overuse passive constructions*. But keep in mind that the subject of the sentence determines whether active or passive voice is required. The most important thing is to make sure that the sentence subject is the same as the sentence topic.
17. *Use the past tense when describing how you performed your study and what you found or what other researchers did; use the present tense for general statements and interpretations.*
18. *Make figures and tables easy to understand without a need for reference to the main body of the article*. Omit data that are not informative. In captions or footnotes of figures, define all abbreviations and symbols that are not obvious. Use text tables when presenting a small set of data.
19. *Define abbreviations when they first appear in the main body of the text.* Avoid abbreviations in the abstract.
20. *Do not write about yourself as "the author(s)"*, as this is ambiguous. Instead, write "we" or "I". More and more journals prefer this style.
21. *Ask a thoughtful colleague to read the whole text, in order to see if anything is ambiguous or unclear.*

These suggestions are based on a range of editorial recommendations for authors and translators of scientific articles. *"If authors and translators follow these guidelines before submission, their manuscripts will be more likely to be accepted for publication."* (EASE)

3.6 Coping with stress

"Life is like riding a bicycle. To keep your balance you must keep moving."[14]

At the end of this chapter we feel the need to discuss an issue that is highly relevant to successful production of science: managing stress. Stress is a state of critical mental (emotional) and physical disorder or imbalance that can lead to more serious medical and psychological complications (headaches, anxiety,

[14] Albert Einstein as quoted in Walter Isaacson, *"Einstein: His Life and Universe"* (2007), p. 367.

sleep disturbances, RSI[15]). Stress may be due to a number of causes. In the case of research work, these are (BLOOMFIELD and EL-FAKAHANY, 2008):

- too many parallel tasks
- too much routine work
- tension in your professional network (unclear roles and responsibilities)
- deadlines
- pressure to compete for funding
- pressure to publish.

Each of these causes can be dealt with by adopting a systematic strategy. For example, too many parallel tasks probably means that you have to learn how to drop out of some collaboration, or limit your tasks to an agreed list. To avoid tension in a group it's often a good idea to increase the frequency of meetings and discussion panels. Spend more time communicating and giving each other a chance to meet and debate. Do not avoid confrontation. As with any project, it's better to have a team that communicates honestly than to pretend that problems do not exist. Likewise, do not hide problems that you cannot explain — seek help. Deadlines cannot be avoided, but at least you can prepare yourself psychologically (e.g. if you are a graduate student take a look at Fig. 5.5 to know what to expect). Think of a deadline as like an important game: the further you get in the playoffs, the more serious you need to be. Now imagine a positive outcome (victory). This thought can carry you through the tough times.

You can deal with stress and tension in your professional network by improving your social practice in general. For example, here are some general suggestions on how to improve chemistry with your colleagues:

- try meeting your colleagues in an informal setting (e.g. in so-called *"team building"* sessions)
- try participating in group sports
- visit other research groups and learn from their experiences
- visit research groups abroad (e.g. on sabbatical) — observe how things are organized and what is better (or worse) compared to your own organization
- attend summer schools or workshops in a less formal setting
- attend conferences that focus on new developments, new techniques and fresh ideas
- attend workshops that stimulate brainstorming and interdisciplinary exercises — *"games are the most elevated form of investigation."*[16]

Socializing in science is good. Creative individuals inspire each other and great ideas come out of interactive brainstorming — these people are often your best co-authors. The problem is to find them.

Many inexperienced researchers work too hard. The number of hours spent working is, in fact, irrelevant. It's the quality of the scientific work that counts

[15] Repetitive Strain Injury — damage to the musculoskeletal and/or nervous system that may be caused by repetitive tasks.

[16] Quote by Albert Einstein.

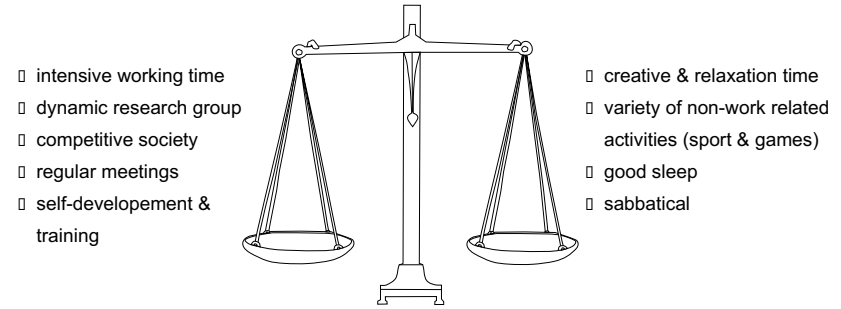


Fig. 3.3 The key to success in academic work is a balanced combination of hard work and creative relaxation time.

(ASCHERON and KICKUTH, 2004). A working environment can play a key role here. If you're not able to work in an environment that allows you to think clearly and, if you don't have information systems that allow you to run analyses, visualize data, and compare your work with that of others, you're unlikely to be able to develop top science even if you are strongly motivated to succeed. Malcolm Gladwell: *"Success is not a random act. It arises out of a predictable and powerful set of circumstances and opportunities."*

Ultimately, you cannot increase your output simply by extending your working hours. After a point, workaholism and obsession become counterproductive. In fact, you can do damage to yourself and your career if you do not find a good balance between creative, relaxing and working time.

Table 3.1 Leo Esaki's rules of thumb for making a career in science. Adapted from ASCHERON and KICKUTH (2004).

Creative side	*Productive side*
• focus on relevant topics only	• generate many publications (acquire significant funds)
• be unconventional and excentric	• be systematic and follow conventions
• be stubborn	• adjust to your research group
• gain independence (of authorities)	• listen to advice from your supervisors
• do not overload your mind with too much information	• memorize important concepts (even computer code)
• sleep on ideas	• publish as soon as possible
• count on luck	• eliminate random effects
• create creative chaos!	• tidy up your desk!

Leo Esaki, the winner of the Nobel Prize in Physics in 1973, suggested a number of practical rules for making a career in science (see Table 3.1). This shows how a researcher has a dual nature — productive and creative — and that

the two are often in conflict. A combination of creative freedom and systematic work is usually the best strategy. The trick is to know in which situations you can improvise and be relaxed, and in which ones you should be painstakingly precise and focused.

Chapter 4
Writing research articles

OK. You have all the data, you're sure about your results and the message that you want to transmit, and you are sure that there is an audience for it — now you can start writing the paper. If all of the above criteria are satisfied, you might think that writing the article would be the easiest part of the job. But it requires considerable skill.

Scientific writing should be easy. Basically, it only requires two tenses: simple present & simple past, so scientific writers only have to master a narrow selection of the rich range of possibilities within the language. However, inexperienced and experienced scientists alike tend to imitate dense nominalized, depersonalized writing styles and stock phrases, such as *"it has been shown that…"* and *"it was observed to…"*. We need to return to some simple ground rules for clear, connected readable style.

Writers of science want a simple *"recipe"* to help them write readable, scientifically credible journal articles. They want to write their articles as quickly as possible and get on with their research. The approach we present below provides three levels of structure that form the basis not only of the article, but also of logically organizing scientific ideas:

- *Macro level*: the overall structure of the article (what goes where)
- *Meso level*: the structure of paragraphs — presenting and supporting scientific messages
- *Micro level*: the structure of sentences — the basic building blocks that tie the whole thing together.

Because information is much easier to interpret if it is placed where most readers expect to find it, many scientific articles follow the broad IMRaD macro structure (Introduction, Method, Results, and Discussion). This provides a kind of road map for readers. In particular, writing the methods and results sections can be pretty straightforward.

You can use the overall structure of the article discussed in this chapter (see also the appendix) as a template for organizing the ideas and messages that you want to publish. In later sections, we will work on building effective paragraphs

(meso level) and sentences (micro level) so that you can clearly present these messages in each section of the article. Then, we will provide some practical advice on how to improve the flow of the paper and increase its readability. Finally, in the appendix you can find a series of steps that will help you organize your article from initial idea to camera-ready-copy.

4.1 Establishing a framework for research papers: a recipe

"The last thing one knows in constructing a work is what to put in first."[1]

You could think of writing a research paper as a bit like constructing a house. We start by putting down foundations and building a frame — its macro-structure. The macro-structure of a manuscript is reflected in the major headings. As mentioned previously, most journals in the world have accepted the **IMRaD** structure as the international standard. For example, the International Committee of Medical Journal Editors promotes use of uniform rules when preparing manuscripts submitted to Biomedical journals — **the Vancouver guidelines**[2]. The Vancouver Rules are also accepted as the basis for publication practice by Committee on Publication Ethics (COPE), Deutsche Forschungsgemeinschaft (DFG), and similar organizations.

Ed Hull's ten-step recipe for making the IMRaD structure work looks like this[3]:

THE INTRODUCTION

1. *Describe a big problem that needs to be solved* — Journal editors reject articles because the point of the research — its relevance — is not immediately and obviously clear. Right at the top of the Introduction show the relevance of your research — present the big problem that your research helps to solve. A word of warning here, presenting a gap in knowledge is not enough. What we need to know are the consequences of that gap in knowledge.
2. *State your strategy to help solve the problem* — This part of the Introduction sharpens the focus on the point of the research. It takes the reader step-by-step from what is already known about the problem to what is unknown about the problem — it logically leads the reader from the "big problem to be solved" to the specific research question.
3. *State a specific research question/hypothesis whose answer/test will help to solve that problem* — This needs to be specifically stated in terms of

[1] After CREEDY (2008)

[2] `http://www.icmje.org/manuscript_1prepare.html`

[3] This outline includes material developed by the experienced writing trainer Ed Hull.

measurable/observable independent and outcome variables and their relationships that your methods were designed to determine. Note that such explicit wording helps the reader to understand your Methods section. A major reason for rejection is that no research question is clearly stated.

METHODS

4. *Describe the methods used to answer that question* — This section will differ depending on the type of research, but 3 main items need to be emphasized: (1) *What was studied*, (2) *How the data was collected/observed* and, (3) *How the data was analyzed to determine relationships between the independent and outcome variables*. This section reports what was done — historical facts — and must therefore be written in the past tense.

RESULTS

5. *Describe your findings* — This section should link directly to the Methods section and emphasize 3 main messages: (1) *The characteristics of the objects you are investigating*, (2) *Results = data that relates to the research question*, (3) *The relationships (e.g. correlations) between the independent and outcome variables that were determined*. Just as in the Methods section, this section reports historical facts — what was found — and must be in the past tense.

DISCUSSION

6. *Answer the research question* — An answer to a research question is a present tense statement of the author's interpretation of his/her results. It links directly back to the research question/hypothesis stated in the Introduction and, therefore, it uses exactly the same words that were used to state the question/hypothesis. As an expert in your field, readers expect you to interpret your results. A word of warning here, do not summarize/repeat your findings in the past tense. Such a repeat is not an interpretation of those findings. Interpretations must be in the present tense. This blunder quite often leads to rejection.
7. *Support that answer* — The author must support the answer to the research question. This can be done in several ways: (1) *by showing how the factual findings, expressed in past tense, support it*, (2) *by relating the findings to the work of others*, (3) by *presenting theoretical considerations that support it*.
8. *State the limitations of that answer* — Every research study has limitations, even yours (I suggest a subheading "Limitations" in the Discussion section). Explicitly state the limitations of your study, and show how they restrict generalization of your answer. Inadequate discussion of limitations is often a reason for rejection.

CONCLUSIONS

(I suggest a subheading "Conclusions" at the end of the Discussion section.) This subsection should clearly state 2 main messages:

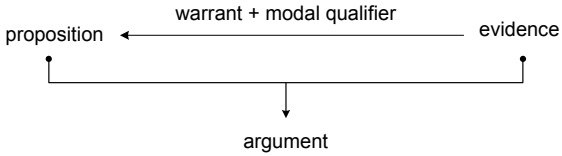
9. *Explain the practical/theoretical consequences of the answer* — Here, you should point out the value of your work. This relates directly back to the problem stated at the beginning of the Introduction. It clearly shows how the work takes a step toward solving that problem.
10. *Propose a next step to help solve the original problem* — One research study seldom solves a big problem. But you — as an expert in your field — need to *"stand above"* the details of your work and tell us a possible next step toward solving the problem. A next step could be: (1) *a new research question*, (2) *a refinement of the present study to reduce limitations*, (3) *a protocol that can be used to implement findings*.

So the key components of a research article are: (1) a significant problem, (2) a strategy for helping to solve it, and (3) possible implications of results for the current state of knowledge. Put these major points on paper and expand your article from this core structure. For more details about the purpose of each section, see the appendix.

4.2 Basic principles of logical reasoning

"No! Scientists do not compromise. Our minds are trained to synthesize facts and come to inarguable conclusions. Not to mention Sheldon is bat-crap crazy."[4]

The basic skill required to produce new information/knowledge is logical reasoning. Especially in the applied sciences, one relies not only on mind experiments, but also tries to back up claims by using evidence — proofs, examples and/or arguments. **Argumentation** is reasoning designed to arrive at the best approximation of truth using proofs — *"a constructive debate to reach a solution"* (ROSSITER, 2009). In the most simple terms, an argument consists of a proposition that is backed up by evidence, connected by some warrant (or justification) with a modal qualifier that expresses the extent of the proposition:



Argumentation in a research paper follows a circular path (Fig. 4.1). We start with an assumption (working claim), then build our case (provide evidence),

[4] Leonard responding to Penny's proposal to make peace with Sheldon; from *The Big Bang Theory* TV series created by Chuck Lorre and Bill Prady.

then make the final claim and justify it with previous arguments. In that sense a research paper is an extensive and detailed version of the argumentation process (TURABIAN, 2007).

Clarity of writing follows clarity of thought. So first think what you want to say. Then write it down as simply as possible.

A research paper can contain several arguments and/or micro-arguments, which are usually interconnected. In principle, every new proposition or claim the authors make in a research paper should be supported by evidence. The evidence could be authors' own data, or other people's data, but a general expectation in any research paper is that we repeatedly present evidence for every new claim we make.

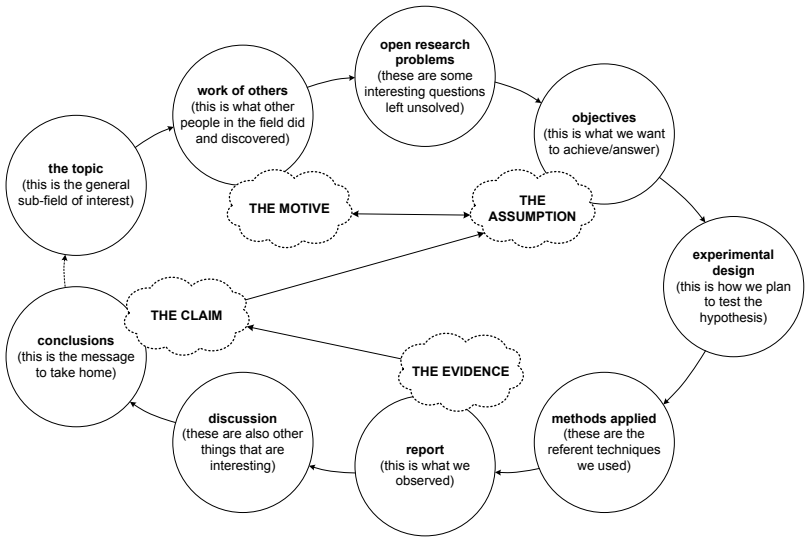


Fig. 4.1 Logical steps in a research paper. Research papers are in essence circular because conclusions refer to claims made at the beginning; likewise, each discovery leads to another cycle of research.

In addition, no claim can ever be absolutely universal, so it needs to include a modal qualifier i.e. it should clarify under which conditions the claim is correct. This issue is further discussed in section 4.7 (page 92). The formal structure of logical reasoning is not as important as the quality of the evidence and order of the reasoning. For some examples of how to build arguments see ROSSITER (2009).

For Edward Tufte[5], a presentation can be much improved if the presenter emphasizes three key elements: *the message* (going from: problem ↦ relevance ↦ solution), *credibility* (why are you the best person to present this work?) and *originality* (e.g. new visualization concepts, new technologies).

Logical reasoning is crucial. An experienced reviewer can easily scan a research paper and detect poor argumentation — usually due to a lack of evidence and/or a claim that is too general. The poorer the arguments, the less credible the paper is. Here are some common causes of flawed arguments (based on ASCHERON and KICKUTH (2004)):

- *ambiguous and vague terms* (your terms are not unequivocal and concrete)
- *wrong evidence* (your results do not match your research question)
- *incomplete evidence* (what is the uncertainty of your evidence?)
- *wrong logical reasoning* (your conclusions do not make sense)
- *unconvincing reasoning* (e.g. your data is not representative of the whole population)
- *missing reasoning* (you do not really know why the results show what they show)
- *exaggeration / bias* (your conclude too much from too little evidence)
- *missing quantification of significance* (how significant is your evidence? could it have happened by chance?)
- *spurious correlation* (you cannot explain the correlation observed)
- *errors in data* (you failed to double-check your results).

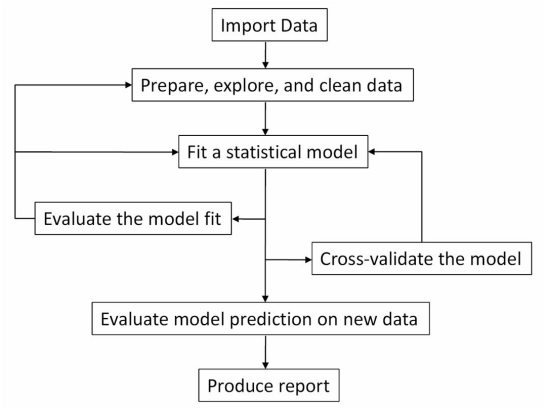


Fig. 4.2 A general data analysis flowchart by KABACOFF (2011).

Logical reasoning is closely connected to statistical reasoning. At the heart of many research papers is some sort of analysis, i.e. the application of robust

[5] `http://www.edwardtufte.com`

analytical techniques to data. For example, ADRIENKO and ADRIENKO (2005) think that any data analysis can be described by the following five steps:

1. *formulate questions*
2. *choose the analysis methods*
3. *prepare the data for analysis*
4. *apply the methods to the data*
5. *interpret and evaluate the results obtained.*

Some generic analysis techniques used to develop evidence are for example:

- disaggregation / analysis (X can be split into X_a, X_b,...),
- aggregation / summaries (sum of X; standard deviation of X),
- correlation / regression analysis (X causes changes in Y),
- time sequence analysis (X decreases after time t),
- analogy / comparison (X is higher with method A than with method B),
- simulation (generate synthetic X assuming model m).

And, of course, there are numerous combinations of these techniques.

Note also that to run analysis on data only once might not be sufficiently convincing. Probability theory teaches us that one *'draw'* is insignificant, because it can accidentally be any value from the probability distribution. We can only convince people of the validity of our claims if we repeatedly prove them using new data. This could be impractical, as the costs of collecting data are typically the most expensive part of research. A technique that makes it possible to re-run analysis over and over again, even on existing data, is **cross-validation**[6]. Cross-validation is now considered to be a standard step in any model evaluation. The challenge is to find a creative way to run it.

4.3 Figures and tables (generating and editing graphics)

"Show, don't tell: seeing is believing."[7]

Figures and tables in an article are your chance to produce original visualizations of concepts or data. An image is said to be worth 1000 words, but to make a high-quality graph or table can take much more time than to write 1000 words. Hence you should start preparing figures in parallel to writing or even before you start writing.

Figures in papers (graphics or artwork) can be of various types:

[6] Cross-validation implies that we randomly split the original data into calibration (model building) and validation (model evaluation) sets to get an unbiased estimate of the model error. Note that it can be repeated many times.

[7] `http://rubyonrails.org/screencasts`

Sketches
Sketches are graphical, simplified representations of rough drawings of important points or key elements of a system.

Flowcharts
Flowcharts visually display processes and their relationships. These can be e.g. processing steps or decision trees.

Statistical plots
Statistical plots are standard plots generated as a result of statistical analysis. The most commonly used statistical plots are correlation plots, trend plots, box plots, and histograms.

Cartographic materials
Any realistic presentation of geographic phenomena is in fact a map, and every map needs to follow basic cartographic principles.

Photographs or images
Many research articles include photographs (visible light) and/or close range or remote sensing images or scans of objects or surfaces.

In principle, each type of graphics requires a special kind of skill and can be evaluated using objective criteria. The worst thing you can do is underestimate the expertise needed to produce high-quality graphics. People study graphic design and typesetting for years, so it's easy to spot the work of an amateur. Maybe the best way to learn how to make convincing graphics is to learn from bad examples. WAINER (1984) provide a number of amusing examples on how to display data badly. Karl W. Broman from the University of Wisconsin-Madison has created a gallery of top 10 *'worst'* graphs (with apologies to the authors), plus a discussion on what's wrong with them and what should have been done in the first instance[8].

Typical examples of poor graphic design are:

(*a*) too many variables on the same plot
(*b*) wrongly emphasized features
(*c*) inappropriate scale
(*d*) too many lines over a small area (the *'spaghetti-effect'*)
(*e*) mixing serif and sans serif fonts within the same graph
(*f*) too small / too large font in relation to the size when printed
(*g*) too thin lines
(*h*) over-compressed images and graphs converted to too low resolution (e.g. $<$150 DPI)
(*i*) missing axis labels and units on graphs or 2D plots.

Figures and tables should be easy to understand without reference to the main body of the article. Graphics that are not informative or those that are

[8] `http://www.biostat.wisc.edu/~kbroman/topten_worstgraphs/`

obvious should be omitted. All abbreviations and symbols, on the other hand, that are not obvious need to be defined in captions or footnotes.

Theoretically speaking, the efficiency of the graphics in a paper can be measured by assessing the **information resolution**, expressed as (TUFTE, 1992):

$$\text{information resolution} = \frac{\text{bits}}{\text{time} \times \text{area}} - \text{noise} \tag{4.1}$$

Although this parameter is often not assessed for all graphics, it is obvious that the quality of graphics will increase if we produce info-dense graphics (i.e. plots or charts that consist of many information elements) that are still readable and clear.

From our experience, producing quality graphics requires adherence to the following guidelines:

- *Sketches* — Sketches are simplified, symbolic representations of ideas, so make sure you use simplified graphical items and nothing too complex. It's not a good idea to mix sketches with objective representations of reality such as statistical plots, because this can cause confusion — sketches by definition do not have to reflect true dimensions and ratios. They are only rough illustrations of concepts.
- *Flowcharts* — A typical mistake people make with flowcharts is that they mix them with sketches. Flowcharts need to be complete, i.e. they should not contain any dead ends or decisions that are ambiguous. Imagine transferring a flowchart to an algorithm. If some parts are not complete, then you will not be able to implement them. Flowcharts can best be produced using specialized software such as Microsoft Visio or Open Office Draw.
- *Statistical plots* — Producing statistical plots is a science in itself. The most common mistake authors make is that they either miss out labels or they mismanage graphical elements. Use decimal points (not decimal commas) for real numbers. We recommend you check out the literature, which is extensive. For example the R community has now several textbooks on multivariate (SARKAR, 2008; WICKHAM, 2009) and interactive graphics (THEUS and URBANEK, 2008). You can also simply browse examples of statistical plots used in R[9,10] and then try to adopt and/or improve them.
- *Cartographic materials* — Maps are often 2D rectangular graphs, but with many specific properties. In order to avoid major misuse of cartographic data, always indicate in the plot or in the text: map source, the effective scale, projection type i.e. coordinate system and map orientation (direction of north). If possible, also indicate mapping accuracy and refer to a URL where there is more technical information (metadata) about the map.
- *Photographs or images* — Photographs and images are typically easier to import into an article. The key issue is that the resolution of such images should correspond to the rule of thumb 150–600 dots per inch (DPI). This

[9] `http://addictedtor.free.fr`

[10] `http://bm2.genes.nig.ac.jp/RGM2/`

means that if the width of an image in the text is e.g. 8 cm, then the size of image should be at least 450 pix, ideally 900 pix (document distribution), and not more than 1800 pix (press). Also be aware that color images will often need to be converted to gray-scale images in journals. Hence it is a good idea to do this conversion before submitting the article. The way to improve the images even more is to maximize the contrast by using some standard histogram equalization functions implemented in image processing packages such as Adobe Illustrator or PaintShop.

The European Association of Science Editors (EASE) emphasizes in their general guidelines[11] that manipulation of images to make a false impression constitutes scientific fraud, so graphical editing of such materials should be taken very seriously.

Tufte believes that the best way to learn to make high-quality graphics is to look at the templates used by Nature, Science or other high-impact journals. People who publish in these journals usually have 1.5 pages of space to tell a complex, significant story. Clearly, the graphics that make it into such articles must be the best. Consider, for example, the artwork and graphics used in The Economist, the New York Times and Science Mag. Everything in these plots is high art, right down to the smallest detail. And nothing stops you from learning from the masters. Tufte: *"Talent imitates, genius steals!"*[12]

Tufte goes beyond standard graphics and suggests the use of so-called *"supergraphics"* — high-resolution visuals with a high density of information per area (take a look for example at Fig. 1.3). Such supergraphics impress and inspire people — they immediately get involved with the topic and start interpreting and analyzing. David McCandless, an information designer from London, has dedicated his career to discovering original ways to visualize important information[13]. It is striking how many hidden connections and patterns there are in data we already know.

Impressive tabular materials are even more difficult to design and use than artwork. Good table design, for example, is extremely difficult. MS Office or similar text editors will not be of much help. Likewise, journals typically do not provide table templates that you can easily adapt for your data. LaTeX, for example, has several packages that allow production of super-tables (tables that spread over several pages) and allows combination formulas, plots, numbers and text tables[14]. By combining LaTeX and R one can produce high-quality scientific graphics that will impress both scientists and wider audiences.

Many commercial publishers run automatic checks for Artwork Quality. This, for example, is the result of an artwork quality review of a plot that was produced using R and then exported into PostScript format:

---

[11] `http://www.ease.org.uk/guidelines/`

[12] Edward Tufte at a one-day workshop on presenting data and information held on March 28, 2010, Pittsburg (PA).

[13] `http://www.informationisbeautiful.net`

[14] `http://www.macrotex.net/texbooks/`

Help to Improve Image:	http://www.elsevier.com/artwork

Image Details

Resolution		File format	
Dpi (Dots per inch):	300x300	File Type:	EPS / PostScript
Image Size (inches):	8.4433x4.1233	Colours:	2
Print size (cm at 300 dpi):	21.45x10.47	Usable for Review?:	Yes
Usable:	Yes	Usable for Production?:	Yes
Comment:	This vector file is probably usable	Comment:	We will endeavour to use this file in its native format.

Print Close

In order to see whether your graphics will really look good in your article you should consider building a PDF with formatting close to the Camera Ready Copy formatting used by the publisher. For example, while writing this book we kept adjusting sizes and widths in LaTeX, which allowed us to find an optimal combination of text and graphics. There are also commercial packages that you can use, such as Adobe PageMaker, Microsoft Publisher and Scientific WorkPlace.

4.4 Citing sources

> *"The purpose of academic citation is to give credit to those whose work and ideas have shaped your thought, to clearly differentiate between what you have done and what others have done, and to help readers locate the texts/work that has influenced you."*[15]

Citing literature sources is as important a skill as any other writing skill. The way we cite and what we cite in our work is becoming ever more important because citation statistics are likely to become THE evaluation criterion.

According to ROSSITER (2009), an author should aim to cite *the most relevant*, *reliable*, and *accessible* sources only — *"Superfluous references do not impress; they confuse"* (ROSSITER, 2009, p.32). Imagine, for example, that you are using a statistical method that was described 50–60 years ago. Imagine that this method has been published in several formats (e.g. as a part of a PhD thesis, technical report, journal article and book section). Since then it has been reviewed and revisited by several other authors. Then, a summary of the method has been added to an encyclopedia and a website that specializes in this topic. There could many relevant references for this method, so we can easily get lost in figuring out which one to choose (should we cite them all?).

[15] Adapted from the e-mail forum of the Middle East and North Africa Writing Centers Association: `http://menawca.org`.

According to ROSSITER (2009), there is a logical sequence that everyone should follow when citing sources, approximately:

- First, try to use the most original peer-reviewed reference published by an international publisher.
- If this is not available, then refer to the most original peer-reviewed publication even if it produced is by some local or regional publisher (e.g. via a non-ISI journal).
- If this is not available, then refer to the most original publication from a publisher or academic institution, one that has been catalogued via the WorldCat (i.e. which at least has an ISBN and/or a catalogue number).
- If the only source is unedited conference proceedings, refer to the volume title, conference location and/or hosting institution and URL.
- If this is not available, then at least refer to the most original publication that is accessible via an URL (e.g. an on-line report or a PDF).
- If no other publication is available, but only a cross-reference to a review article, encyclopedia article or newspaper article, then refer to the document so that anybody can find the same source within a period of at least the next five years (if necessary put a PDF online and refer to a permanent URL).

Referring to encyclopedias or lectures notes is not acceptable to most journals. Encyclopedias are only reviews or resumes of topics and hence by definition are not meant to provide new information. If they leave out a reference to the original publication (which is very common), you might get a wrong impression that this is THE source. It rarely is. Likewise, lecture notes are usually written for internal purposes and often create the wrong impression (that the authors are producing something original).

The purpose of academic citation is to give credit where it is due, place the paper in the network of papers and researchers on the topic, and provide documentary evidence for the statements made in your paper.

Many academics rush to put things on the web without a second thought. Big IT companies such as Google or Microsoft *crawl* over www and archive all the content that appears on-line. This content is then available via their cache, which is still on-line even if you decide to remove your PDF from the web[16]. Once you put things on the web for a few days and allow public access — it stays there forever, so think twice before you click that upload button!

Many materials finish up on the web without much filtering and cross-checking. As a rule of thumb, at least half of the content you find on the web contains errors, is outdated or inaccurate (sometimes even on purpose). Commercial publishers, public agencies and academic institutions can at least certify with their name that the published content has gone through some kind of

[16] You can remove some of the Google cached content by logging in to the `http://www.google.com/webmasters/tools/removals`, but this is not a trivial issue.

filtering process, although there is almost never any warranty that the information in publications is 100% correct (even nicely formatted articles can contain a lot of nonsense).

Here's a list of some common misconceptions about who and how to cite:

- Citing literature is in general a good thing[17], but putting in hundreds of irrelevant citations is counter-productive. SMITH (1990), for example, suggests that authors should limit their list to contemporary, relevant and essential references only: *"The purpose of academic citation isn't to show that you are a 'good student' who has read everything on a subject and can't see the forest for the trees."*
- Citing *sexy* papers doesn't mean that yours will become one.
- Citing the work of reviewers (or supervisors) can please them, but if they are fair reviewers than they will not be impressed. If their vanity is overwhelming, then you should maybe rethink publishing in that journal.
- Although it seems easier to cite review articles than to read all of the original publications (it saves you a lot of homework), you should always try to dig into the literature — detective work can be fun!

Ideally, you should refer not only to publication records, but also to page or table/figure numbers. Imagine that someone really wants to check your numbers or claims — wouldn't you like to help them find the source quickly? Imagine you refer to some results (numbers, plots) published in a thick monograph. How much time would it take to find that specific fact? Hours? Days? In fact, it's highly probable that in the near future publishers will cross-link the references to URL repositories of documents. We will then be able to refer to *micro-location* of a cited source. For example, something like this:

- Books $\mapsto$ `ISBN:Page:Line`
- Journal articles $\mapsto$ `DOI:Page:Line`
- Newspapers and daily journals $\mapsto$ `ISSN:Date:Page:Line`
- Video materials $\mapsto$ `OCLC:Minute`
- Web-materials $\mapsto$ `IP:Anchor:Date`

Google has been pushing a project[18] to scan all (!) of the books in the world and index them (which they have done efficiently with many other types of information) so that a micro-reference to a page in a book or article would take you directly to the specific section or paragraph. References will then probably become redundant and we will only need to refer to a unique ID (Fig. 4.3). In the meantime, it might be a good idea to scan pages with important reference sources and share them among your co-authors, just to improve the collaborative writing process.

[17] Some investigations have shown that papers with more references have a slightly higher impact.

[18] `http://books.google.com/googlebooks/library.html`; other similar competing projects are `http://europeana.eu`: the European archive of digital libraries, and `http://www.hathitrust.org`: the digital archive of USA University libraries.

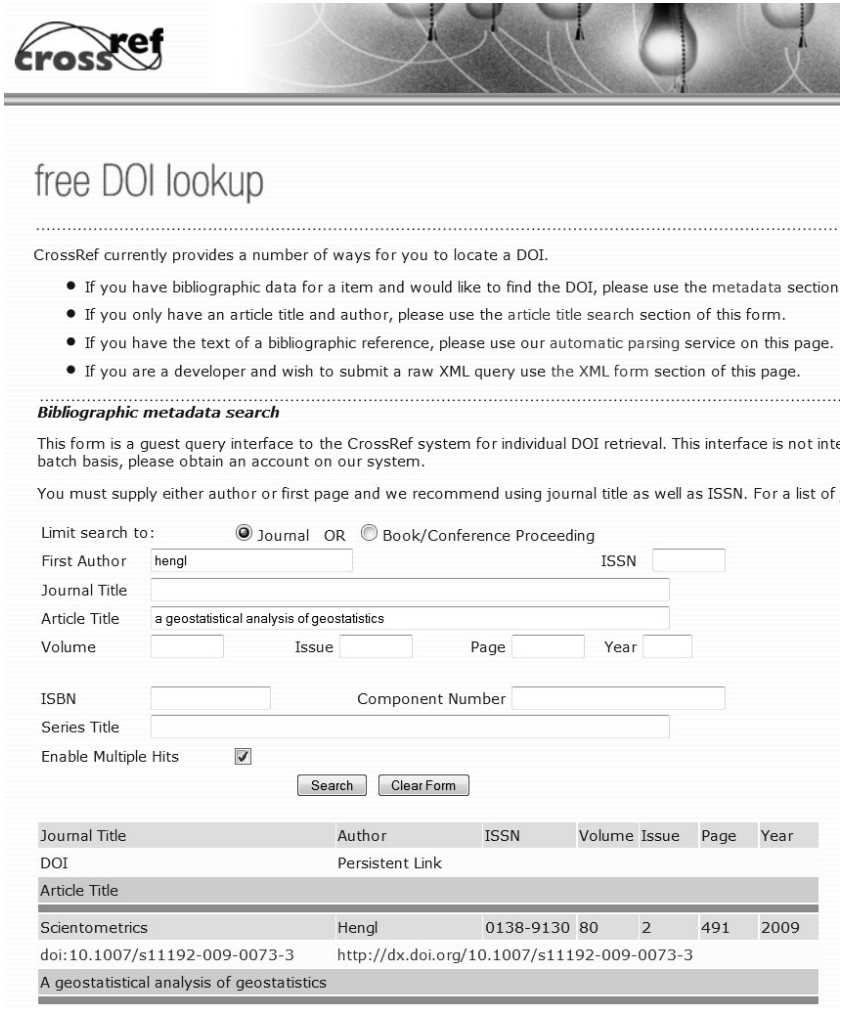


Fig. 4.3 CrossRef.org portal provides a free DOI lookup service where you can find unique identifiers using first author and title. Once you have a DOI or ISBN, you can easily import a reference to your document (BibTeX or EndNote format) and avoid any additional work and typing errors.

References tells a lot about the article — the way we cite and the length of references determines the quality of the research we produce. Here two rules apply. First, read more and you will start to cite more and consequently minimize the chance that you will miss an important reference. Remember that articles that cite more references are in turn cited (slightly) more themselves (WEBSTER *et al.*, 2009). Second, be critical about what you read. If in doubt,

cite only the work that is relevant, accurate and significant, or as CORBYN (2010) puts it: *"to be the best, cite the best"*.

4.5 The art of line editing

> *"Learn to enjoy the tidying process. I don't like to write; I like to have written. But I love to rewrite. I especially like to cut: to press the DELETE key and see an unnecessary word or phrase or sentence vanish into the electricity… with every small refinement I feel that I'm coming nearer to where I would like to arrive, and when I finally get there I know that it was the rewriting, not the writing, that won the game."*[19]

Once you are sure that you and your co-authors are satisfied with the macro (sections) and meso (paragraphs) structure of your article, it's time to zoom further into the paper and analyze the article sentence by sentence. The key to improving the microstructure of a paper is to compress sentences, improve the balance of foreground and background information, and make links between short, *staccato* sentences. This is *"the art of line editing"*.

Here are some general strategies:

- *Compress sentences* by reducing unnecessary or redundant words. Instead of writing *"a considerable number of"* use *"many"*; instead of *"this result would seem to indicate"* use *"this indicates*". Omit obvious sentences such as: *"It is well known that…"* and *"The correlation plot is shown in Fig…"*. However, be careful not to overdo the use of short sentences — JASPER (2002): *"Every sentence cannot be urgent. Good writing normally requires a combination of longer and shorter sentences, carefully orchestrated in each paragraph."*
- *Split long, heavy sentence* into two or even three. Keep paragraphs short and eye-catching. Zinsser: *"Short paragraphs put air around what you write and make it look inviting."*
- *Improve the flow*: at the beginning of a sentence put old information that links back to the previous one and, at the end, put the new information you want readers to focus on. In English, the steps in an argument tend to be built up progressively in coherent paragraphs, with each step linked to the one before, usually through sentence subjects. Example: *"The use of land, water and minerals has increased more than tenfold during the past two centuries* $\mapsto$ *Future increases in population and economic development will intensify this pressure* $\mapsto$ *The cumulative impacts of human activities are likely to lead to major environmental changes, varying from disruption of local ecosystems to disturbance of the biosphere."*
- *Use linking words* such as *although*, *as is clear from*, *as a result of which*, *but*, *most of which*, *or*, *so*, gerund forms (*using… requiring*), *while*, *which is why*, *which indicates that*, *when*, *where*, and *yet*. This approach improves

[19] Willem Zinsser in *"On Writing Well"*.

'message management' by arranging information to show what is important, grouping related ideas, and highlighting the relationships between different parts of the argument. It enhances readability, allows variation in sentence length, and avoids redundancy.

- If too much information is placed at the front of the sentence (*frontal overload*) and too little at the end, the rhythm of the sentence is disturbed and it becomes more difficult for many readers to process the information. Instead of writing: *"Working with students is what attracts me most in this job"* write: *"What attracts me most in this job is working with students."*
- Putting *information in the middle* of a sentence may disturb the flow and can lead to the wrong elements (e.g. pronouns) being stressed, or to redundancy. English readers are unused to the verb coming at the end of a sentence. They process information as they go along. For example, avoid writing: *"It is for the purpose of the present study convenient to… "*.

As GOPEN (2004) points out in *"Expectations: Teaching Writing from the Reader's Perspective"*, when trying to understand a sentence, readers need to find the answers to five important questions:

1. *What's going on here?*
2. *Whose story is it?*
3. *What is the most important piece of information in this sentence?*
4. *How does this sentence link backwards to the one that precedes it?*
5. *How does this sentence lean forwards to the one that follows it?*

Cutting things out

One efficient way to improve coherence is to re-read paragraphs from the beginning and then:

- reorganize and inserting missing links between paragraphs where necessary
- delete redundant content
- improve the connection between sentences (all sentences within a paragraph should be logically connected; otherwise make new paragraphs)
- compress long, over-complex, ambiguous sentences using normal everyday language.

In his general guide to writing, ZINSSER (2006) suggests that in most drafts 50% of the content can be cut out without losing the author's voice. ZINSSER (2006) further suggests that the key to improving the value of a manuscript is in decreasing the *'clutter'*[20] — *"fighting clutter is like fighting weeds"*. To cut things out, however, requires skill. Zinsser: *"If you give me an eight-page article and I tell you to cut it to four pages, you'll howl and say it can't be done. Then you'll go home and do it, and it will be much better. After that comes the hard part: cutting it to three."*

[20] According to the Cluetrain Manifesto (`http://www.cluetrain.com`), *"business babble"* is the types of language used by corporations to hide their mistakes.

If in doubt, cut it out. CREEDY (2008) puts it like this: *"Some well-chosen cuts can reduce the length while at the same time improving the flow of a sentence."* Some editors suggest that in most papers submitted to journals, it's possible to omit the first paragraph without serious loss of meaning (CREEDY, 2008). There's no need to tell your audience things they already know. There's also no need to use empty verbs (see examples below) that overload sentences and irritate readers. A must read for anyone that wants to develop skill in reducing noise in your documents is DUPRÉ (1998).

Correct specialist terminology is clearly essential in scientific writing, but you should use the words and patterns of normal, everyday language to link scientific terms and concepts. Beware of pseudo-scientific *'empty'* verbs. Below are some examples[21] of correct and incorrect uses of *"observed"* and *"showed"*:

Potentially empty verbs: *'observed'*

1. Original sentence (from a review):

   "These differences are for a large part caused by the different modeling methods observed, with differences in time horizon as one of the most prominent causes."

   Is *'observed'* necessary here? No.

   "These differences were caused largely by different modeling methods, with differences in time horizon as one of the most prominent causes."

2. Original sentence:

   "Apart from these essential differences on time horizon, large differences between modeling methods were observed in general."

   Is 'observed' necessary here? No. Authors have an opinion, but this is not connected to the actual measurements.

   "As well as these essential differences in terms of the time horizon, there were large overall differences between modeling methods."

3. Original sentence:

   "Tumor regression was observed in 16 selected patients during curative radiotherapy. Patient selection was based on visual tumor regression on cone-beam CT scan without atelectasis."

   Is 'observed' necessary here? Not if the writer means this:

   "Tumors regressed in 16 selected patients during curative radiotherapy. Patient selection was based on visual tumor regression on cone-beam CT scan without atelectasis."

   But she actually meant this:

[21] These examples were collected by academic writing teacher David Alexander.

> *"To observe (the process of) tumor regression during radiotherapy, we selected 16 patients on the basis of visual tumor regression on cone-beam CT scan without atelectasis."*

4. Original sentence:

> *"This raises questions with regard to microscopic pathological response, which cannot be observed on CT-scans."*

Is 'observed' necessary here? Yes.

Potentially empty verbs: *'show'*

1. Original sentence:

> *"The model showed that the throughput time of the diagnostic track could be reduced from 21 to 5 working days with 18 mins of idle time per day & 10 minutes of overtime per day."*

Is *'showed'* necessary here? Yes — the model demonstrated something, which is what models are designed to do.

2. Original sentence:

> *"We show that neither the combined expression of TTFCE6SH and TTFCE7SH nor the combination of one of the shuffled genes with its WT counterpart enabled transfected NIH 3T3 cells to form colonies in soft agar."*

Is *'show'* necessary here? Yes, science readers expect other researchers to show them things.

3. Original sentence:

> *"The best results were obtained with a 1/9/1 w/w ternary mixture of docetaxel, polyvinylpyrrolidone (PVP)-K30 and sodium dodecyl sulphate (SDS) which showed a significant increase in rate and extent of in-vitro dissolution of docetaxel."*

Is *'showed'* necessary here? Not if the writer means to tell us something like this:

> *"The best results were obtained with a 1/9/1 w/w ternary mixture of docetaxel, polyvinylpyrrolidone (PVP)-K30 and sodium dodecyl sulphate (SDS), which significantly increased the rate and extent of in-vitro dissolution of docetaxel."*

Before writing *'showed'*, ask yourself what actually happened. Here, the main effect of the mixture was to increase something. Much in the same way as genes don't *'show'* expression, but are expressed, mixtures don't usually *'show'* anybody anything.

4. Original sentence:

> *"The RING domain structure resembles other U-box/RING domains, but the N-terminal helixes show an interesting asymmetric arrangement, a phenomenon that has not previously been seen in such domains."*

Is 'showed' necessary here? No. *"Are arranged"* is a much more suitable verb:

> *"The RING domain structure resembles other U–box/RING domains, but the N–terminal helixes are arranged asymmetrically, a phenomenon that has not previously been seen in such domains."*

Making strong(er) points

At the micro-level, you might want to strengthen (or weaken) your arguments. If you are not a native English speaker, consider using some of the following phrases to build effective arguments:

- the objective of this study is to draw a distinction between / make a comparison between…
- our aim is to raise important questions / take into consideration…
- we want to make a case for / put greater emphasis on…
- many studies have attempted to assess the significance of…
- in study we wish to draw attention to new research which suggests…
- here we present the case for… / we put forward the argument that…
- we draw the conclusion that…

On the use of *however*, *therefore*, *so* and *thus* — In research on what makes texts readable or not scholarly writing ranks the lowest, not just for the content but also for the *'style'* in which it's written (RAYMOND, 1993). A number of textual factors are analyzed: length of sentences, word choice, etc. One factor is the use and placement of linking words such as *'however'* (as well as thus, therefore, and, but, also, because, and although). Placing these words at the beginning of sentences can help to signal to readers quickly — without them having to delve — when the author intends to make a logical connection or shift. However, if it sounds unnatural, it will reduce readability, so the impact may be different on native and non-native speakers of English. The use of *'however'* and *'thus'* at the beginning of sentences is different in written and spoken English. And written English in the US tends more than British English to mimic spoken English. Why can the connector *'so'* be placed in the initial position, while other connectors are less preferred in that position? Clearly, there is no actual grammatical rule forbidding connectors at the beginning of sentences. And native speakers do use them. So it's not actually wrong. The difference with *'so'* is that a sentence starting with *'so'* flows along quite happily, whereas one that starts with *'thus'* has to come to a stop before it gets started.

The following sentences make somewhat weaker arguments:

- The data broadly supports the view that…
- It appears that the effects of … can be explained with…
- We provide evidence to support the claim that…
- The results indicate that there is a connection between…

Some scientists tend to *'hedge'* i.e. use weaker arguments that are justified by their research. In such cases, take a stronger stand and be more assertive about your discoveries and conclusions e.g.:

- The results of this study confirm that…
- This is a clear illustration of…
- The data does (or does not?) show… and this trend is significant.
- We offer proof that…
- We challenge the theory…

Of course you don't only need to choose your arguments well. You also need to select facts to focus on. Remember that the most interesting facts are:

1. *simple*
2. *analogous to many other facts*
3. *grouped with others (i.e. not isolated)*
4. *those that have a greater chance of recurring*.

How do you choose which facts to emphasize? Select those that possess beauty and harmony. The more general a rule is, the greater its value. The best rules are simple and elegant. Once a rule is established, the exceptions become important. But be prepared to break rules where necessary:

> *"Life is short, Break the rules, Forgive quickly, Kiss slowly, Love truly, Laugh uncontrollably, And never regret anything that made you smile."*[22]

4.6 Common language mistakes

> *"I would never use a long word where a short one would answer the purpose. I know there are professors in this country who 'ligate' arteries. Other surgeons tie them, and it stops the bleeding just as well."*[23]

Editors have discovered that many inexperienced authors (especially non-native English speakers) often repeat the same mistakes. Here are some common ones:

- *Active or passive?* — Note that it may be necessary to use a passive verb to maintain sentence flow. However, wherever possible, use active constructions within sentences (e.g. write *"This programme focuses on…"*, not *"The programme is focused on…"*). *"The difference between an active verb style and a passive-verb style — in clarity and vigor — is the difference between life and death for a writer"* (ZINSSER, 2006).
- *He or she?* — Use plural forms to avoid sexist writing. Instead of *"…this affects the end-user, unless 'he' possesses…"* $\mapsto$ *"end-users…unless 'they' possess…"*. If this solution doesn't work, use *"he or she"* / *"his or her"*.
- *Color or colour?* — Use either UK or US spelling and use it consistently. Follow the spelling in the Oxford Advanced Learner's Dictionary of English (for instance: analyse, honour, colour, realize, organize, programme, centre)

[22] Combination of quotes by Robert Doisneau and Amber Deckers.

[23] Oliver Wendell Holmes, Sr.

or use the US Websters dictionary. Words in a title or a heading should be written in lower case, except for the first word and any proper nouns such as names of persons, organizations or countries, if you use UK style. In US heading style all keywords are capitalized.

- *To measure or to make a measurement of?* — Wherever possible use verbs, not heavy, abstract nouns (e.g. write *"reduce*" rather than *"achieve a reduction"*). Nouns make sentences stand still. Verbs make them move and push our meaning across to readers. Turning verbs into nouns hides their action. Use *"adapting to"* instead of *"the adaptation to"*; *"for measuring"* instead of *"for the measurement of"*; a project *"designed to develop"* instead of *"aiming at the development of . . . "*.
- *Tripoli or Tarablus* — Foreign terms for which there are no widely accepted English equivalents should be used with an English translation (in brackets) the first time they appear. Spell all foreign words correctly and pay special attention to diacritical accent marks such as è, é, ä, ö and ü.
- *NGO or ngo?* — Abbreviations must be capitalized. If they can be made plural, this should be done by adding a lower-case "s" without an apostrophe. For instance: NGOs rather than ngo's. Spell out terms that are subsequently abbreviated when you use them for the first time, with the abbreviation between parentheses. For instance: Organization for Economic Cooperation and Development (OECD).
- *Europe or europe?* — Geographical terms commonly accepted as proper names are capitalized. Terms are not capitalized when they denote simply direction or compass points. For instance: The Middle East, but western, central and eastern Europe. For the names of newspapers, periodicals and organizations, official titles, and the like use the original spellings (with English translations in parentheses where necessary).
- *Hyphenation or hy-phe-na-ti-on?* — Hyphenation should always break at syllable boundaries. In reports, leave a blank line between paragraphs and do not indent at the beginning of a new paragraph. Avoid a single line of text at the top or bottom of a page (known as 'widows' and 'orphans').
- *Dot or comma?* — Use the comma (,) rather than the full stop (.) in numbers containing more than four digits (for example: 10,000). Be consistent in your use of currency symbols. These symbols precede the amount of money. Use the euro sign (€); do not use the words euro or euros. Other currencies should be treated similarly. If you need to refer to American dollars, use US$ or USD. Don't leave a space between the currency sign and the amount. For instance: €10,000 rather than € 10,000 (or 10.000 euros). Use a full stop rather than a comma for decimal places (e.g. write 4.25, not 4,25). Round currency figures off to the nearest euro (or other currency). For instance: €10,234.59 must be written as €10,235. Numbers below 10 are usually written in full when used in running text (one, two, three. . . nine) unless the sentence contains a combination of numbers. For instance: *"Nine delegates attended the meeting. The 9 delegates represented 18 organizations*."

- *A small number of or a few?* — Avoid using long-winded expression: find a shorter alternative, e.g. replace *"A small number of"* with *"Few"*; *"Despite the fact that"* with *"Although"*; *"In order to"* with *"To"*; *"Has been engaged in a study of"* with *"Has studied"*; *"There is a reason to believe"* or *"It appears that…"* with *"I think"*

What makes papers look bad? It's either bad logic, bad structure, poor graphics (or a complete lack of them), unclear sentences, vague statements, ambiguity… some papers are too long, some are too short; very often they need to be redrafted.

On the use of 'I' — Much discussion has focused on the use of first person in scientific writing (RAYMOND, 1993). Indeed, many editors do not accept any use of first person style, but an increasing number do (KIRKMAN, 2001, 2004). A traditional, technical (depersonalized) writing style is still a prerequisite for submitting articles to some journals. Many scientists feel it sounds egoistic to write in the first person. Recall the rules of science from page 5 — a research article should be objective and unbiased. However, research articles should not be just a chronological narration of work done (see also the EASE guidelines on page 65). If an author has produced some results, and if these are accurate, then we should all come to the same conclusions and hence the use of *"It can be concluded"* is completely legitimate. However, if you write *"It is assumed that…"*, *"It was decided to…"*, *"Sites were chosen…"*, the reader might ask: who assumes?, who decided? — the author? or his boss? his client? or… Obviously, if such information is missing, readers may completely misunderstand the paper (WEBSTER, 2003). There are situations when authors need to make a clear distinction between their opinion and the opinion/results of others (KIRKMAN, 2001).

4.7 Can you please be more specific?

"When a man sits with a pretty girl for an hour, it seems like a minute. But let him sit on a hot stove for a minute and it's longer than any hour. That's relativity."[24]

Here is an example from the movie Phenomenon[25] in which an FBI agent runs some intelligent tests on Mr Malley (the 'phenomenon' character):

► FBI agent: *"Answer as quickly as you can. How old is a person born in 1928?"*

[24] Quote from the abstract of a short paper written by Einstein that appeared in the now defunct Journal of Exothermic Science and Technology (JEST, Vol. 1, No. 9; 1938).

[25] Written by Gerald Di Pego; distributed by Touchstone Pictures.

- Mr Malley: *"Man or woman?"*
- FBI agent: *"Why?"*
- Mr Malley: *"Specifics, Bob."*
- FBI agent: *"Okay, one more time. How old is a man born in 1928?"*
- Mr Malley: *"Still alive?"*
- FBI agent: *"If a man is born in 1928, and he's still alive, how old is he?"*
- Mr Malley: *"What month?"*
- FBI agent: *"If a man was born October 3, 1928, and he's still alive, how old is he?"*
- Mr Malley: *"What time?"*
- FBI agent: *"10:00."*
- Mr Malley: *"Where?"*
- FBI agent: *"Anywhere!"*
- Mr Malley: *"Well, let's get specific, Bob. I mean, if the guy's still alive, born in California, October 3, 1928, 10:00 p.m., he's 67 years, 9 months, 22 days, 14 hours and ... and 12 minutes. If he was born in New York, he's three hours older now, isn't he?"*
- FBI agent: *"How do you do that?"*

In this example, Mr Malley asks for more detail because his model of time is highly accurate (minutes) so that a specific location on Earth determines someone's precise age. Researchers also prefer to know such details. From the point of view of spatio-temporal statistics, every variable we measure refers to some space-time *'location'*:

1. *geographic location* (longitude and latitude or projected X, Y coordinates)
2. *height above the ground surface* (elevation)
3. *time of measurement* (year, month, day, hour, minute)
4. *spatio-temporal support* (size of the blocks of material associated with measurements; time interval of measurement).

Methods developed to analyze such data require spatio-temporal references to be accurately specified; otherwise no accurate interpretation of the results can be made. Statisticians recommend that any measured variable in environmental sciences should indicate: geographic coordinates (location), applicable vertical dimension[26], a time reference (time interval of measurement), and the size of the sampling blocks (support size).

Take the example of currency. To say that something costs 100 dollars is certainly ambiguous. First, which dollars are you referring to? American, Canadian, Australian... The currency of any dollar is subject to daily fluctuations. For example, 1 US$ was worth the 0.60 EUR in June 2009, but six months later it was worth 0.825 EUR, and in August 2011 it was worth 0.70 EUR[27].

Another issue connected with being more specific about data and results is the applicability scale. Researchers are usually biased in the sense that they

[26] Orthogonal distance from the surface of the ground.

[27] See `http://www.google.com/finance?q=EURUSD`

take a small sample of the whole population and then, based on such limited results, try to build universal theories. According to SIEGFRIED (2010), statistical tests are widely misunderstood and frequently misinterpreted. As a result, countless conclusions in the scientific literature are erroneous, contradictory and/or confusing. To avoid such criticism of your paper you should:

- *Always make conclusions that relate specifically to your results.*
- *Do not claim new knowledge without evidence.*
- *Always refer to statistical significance* and specify: (1) how representative is your sample? and (2) how significant is the difference from random effects?
- *Always indicate the space-time reference* to which the results are applicable.
- *Always assume that measurements are 'noisy', that uncertainty gets magnified, and that results may be 'positive' because of human desire to produce 'positive results'.*

Unfortunately, we're all biased when it comes to our own work and this bias is not easy to remove. As CREEDY (2008) puts it: *"Where subjective judgements are involved concerning the value of work, it is possible that the same filter prevents some useful material from being published, or being published in the most appropriate form."* Blaise Pascal: *"It is man's natural sickness to believe that he possesses the Truth."* Donald Kennedy, the editor of Science, also points out that *"scientists often come to believe so strongly in the validity of their theories that they cease to examine them objectively"*[28]. The first step toward reducing bias in one's work is to acknowledge that it exists.

4.8 On styles

> *"Good writing does not come naturally, though most people seem to think it does… Writing is hard work. A clear sentence is no accident. Very few sentences come out right the first time, or even the third time. Remember this in moments of despair. If you find that writing is hard, it's because it IS hard."*[29]

For ZINSSER (2006) four essential spices of good writing are confidence, enjoyment, intention and integrity. Once you incorporate these successfully, people might get so impressed that they start to recognize you based on your writing. Originality is the highest form of intellectual creation.

How do you create your own style of writing? A distinctive writing style often comes through richness of phrases, expressions, and idiosyncratic use of language. Read authors who inspire you, imitate or extend them, then find your own voice. For example, if you're interested in learning how to write in ordinary everyday language that's accessible to both scientists and non-scientists, you could read Richard Dawkin's book *"The selfish gene"* and/or Bill Bryson's

[28] Published in The Economist, 19 May 2001.

[29] William Zinsser in *"On writing well"*.

"A Short History of Nearly Everything". However, avoid the trap of literally copying sentences from other writers — be yourself, be relaxed and confident.

Also consider acquiring some general writing style guides such as the Chicago Manual of Style and/or the Oxford Style Manual. Some useful style manuals are even available free of charge[30]. There is even a manual from the *"For dummies"* series on how to write research papers (WOODS, 2002). It should not surprise you that these books are thick volumes because scientific writing is rather exact.

The best writers in the world are also the best readers. It is even better if you do some peer review work yourself. Refereeing is a good way to learn to write better papers. SMITH (1990): *"evaluating the work of others gives one insight into one's own."*

Another thing you might consider is to think up original expressions that epitomize your work (these are known as *streamers* or *callouts*). One day, somebody might remember you for this.

Why is it important to write science using everyday language? — For many critical readers, academic writing is a stereotyped joke — a museum of passive voice, obscure and dull jargon, and stilted paragraphs — without excitement or entanglement (JASPER, 2002; WEBSTER, 2003). Early in their careers, most academics develop the idea that editors are impressed by a *'dense'* writing style with heavy abstract nouns instead of lively active verbs. Researchers should use specialist terminology, but link it with everyday language to captivate an audience that is intelligent but does not necessarily know much about the topic. We recommend using the language of everyday speech — not that of spokesmen, lawyers and bureaucrats — to link scientific concepts. For example, use words with an Anglo-Saxon base to link scientific terms, which usually have a Latin/French origin. Fortunately, much scientific and technical writing in recent years has moved strongly in the direction of spoken English — shorter sentences and more direct, plain language. Unfortunately, many scientists still erroneously believe they are required to use long, heavy sentences full of abstract nouns.

This is what CHOMSKY *et al.* (2002, pp.45–46) say about how academics write: *"Intellectuals have a problem: they have to justify their existence. Now, there are few things about the world that are understood. Most of the things that are understood, except perhaps in certain areas of physics, can be explained with very simple words and in very short sentences. But if you do that, you don't become famous, you don't get a job, people don't revere your writing. There's a challenge there for intellectuals to take: to take what is rather simple and make it appear to be something very complicated and very profound. Groups of intellectuals interact that way. They speak amongst each other, and the rest of the world is supposed to admire them, treat them with respect, etc. But translate what they are saying into simple language and you'll often find either nothing*

[30] See the MHRA style guide (`http://style.mhra.org.uk`), for writing in the arts and humanities.

at all or truisms, or absurdities.", which is what Jorge Cham's did with his thesaurus of common academic phrases (deciphering academese):

"To the best of the author's knowledge…"	=	"We were really too lazy to do a real literature search"
"Results were found through direct experimentation."	=	"We played around with it until it worked"
"The data agreed quite well with the predicted model."	=	"If you turn the page upside down and squint, it doesn't look too different"
"It should be noted that…"	=	"OK, so my experiments weren't perfect are you happy now??"
"These results suggest that…"	=	"If we take a huge leap in reasoning, we can get more mileage out of our data..."
"Future work will focus on…"	=	"Yes, we know there is a big flaw, but we promise we'll get to it someday."
"...remains an open question"	=	"We have no clue either."

4.9 Collaborative writing

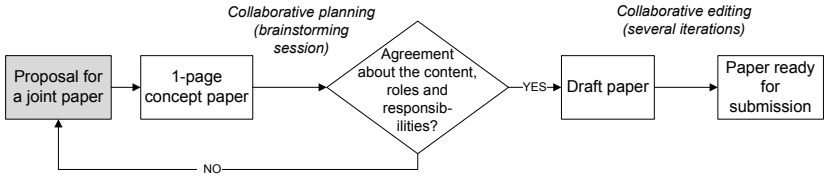
> *"Wikipedia is about the power of people like us to do extraordinary things. People like us write Wikipedia, one word at a time. People like us fund it, one donation at a time. It's proof of our collective potential to change the world."*[31]

Collaborative writing is writing in parallel i.e. more than one person writing in a single copy of a document. Collaborative platforms can be used to generate (1) operational plans (*collaborative planning*), (2) documents (*collaborative writing or editing*), and for (3) annotating text. The advantages of collaborative platforms are:

- They enhance collaboration because the co-authors can have a fruitful discussion without needing to meet physically.
- A significant amount of time can be saved because several people can work on the same document simultaneously. Imagine how much time it takes if you send a document to co-authors by e-mail and then have to wait to receive comments/revisions (from author to author).
- Through collaborative writing co-authors motivate each other to continue writing because the document starts growing and evolving faster.
- The lead author can track contributions from all co-authors and then rank them in the author list.

[31] Jimmy Wales, Founder of Wikipedia.

Collaborative planning (brainstorming; designing) precedes collaborative writing and editing. We recommend that these activities should be separated. In other words, one should not start writing papers in a group before agreeing about roles and responsibilities. **Collaborative editing** comes later. Although some discussion is still welcome at this stage, it is generally better to agree with your co-authors about the main research questions, the target journal, list of authors and their responsibilities (paper production politics) before editing starts.



A collaborative editor is a software application that allows several people to edit a computer file using different computers. Editing can be done both in real time and off-line. In both cases the most important thing about collaborative editing is that the there is a single copy of the document that is accessible by a number of users who all have editing rights and can continuously follow the progress of the paper. Another important feature of collaborative editing is that it allows tracking of changes by author and dates.

Real-time collaborative editing can be best achieved by using web-based editors such as e.g. Google Docs, Adobe Buzzword[32] or a web installation of MediaWiki[33]. There are also a number of desktop packages for collaborative editing, the most popular of which are Microsoft SharePoint, GNU Emacs and Gobby[34].

WIKIPEDIA is the best place to learn about collaborative writing. In fact, it's the ideal place to co-edit and co-share knowledge for a number of reasons:

- It's now the most extensive collaborative encyclopedia in the world, with over 3 million articles.
- It provides a number of tools for writing, editing, translating and exporting content.
- The number of external reviewers is constantly increasing, along with the quality of their feedback.
- Many governments are seriously thinking of using WIKIPEDIA in their educational systems[35], at all levels.

[32] `http://www.buzzword.com`

[33] `http://www.mediawiki.org`

[34] `http://gobby.0x539.de`

[35] The Department for Children, Schools and Families in the UK proposed a draft plan in 2009 for reform of primary education that will *legalize* the use of Twitter and WIKIPEDIA in daily education.

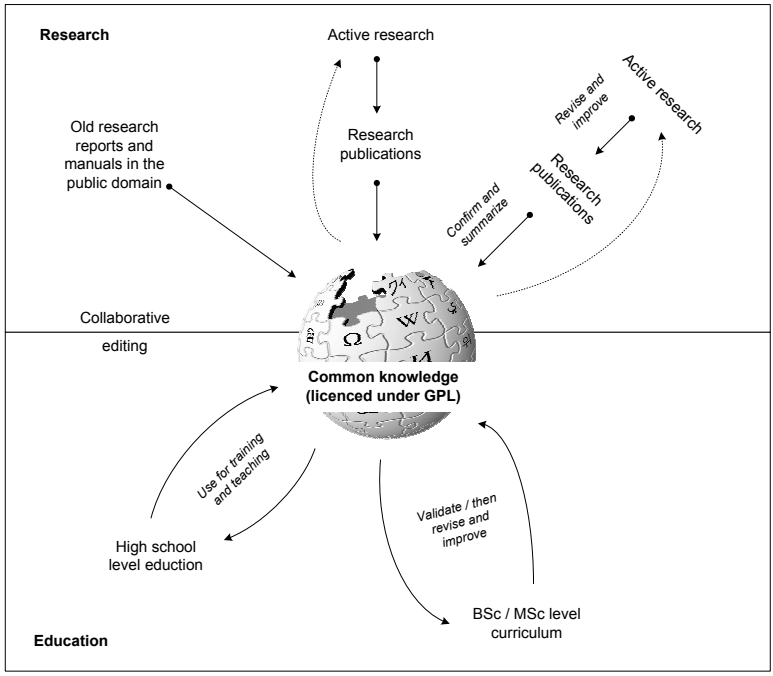


Fig. 4.4 Academic institutions should motivate researchers to contribute content, i.e. the results of their research, to WIKIPEDIA — the public encyclopedia. WIKIPEDIA is among the top five most visited websites in the world, along with Google, Facebook, Youtube,and MSN. The budget of the Wikimedia foundation for 2010 was about 20 million US$; by comparison Google's total costs & expenses for 2010 were approx. 15 billion US$ (or 750 times more).

Why not write everything directly in WIKIPEDIA? Since we are strong supporters of open source software, we should be supporters of free documents such as articles in WIKIPEDIA as well. In order to be free, and to have all the advantages of a true open source project, documents must be made available in editable formats, so that people can contribute to them, change them, correct them… — the same way it's done with open-source software. However, if you write a book in the first person (your own opinions) and based on your own experiences, it would be inappropriate to invite other people to mix their opinions. WIKIPEDIA is by definition not intended for contributions that reflect the results of active research. There is also the issue of evaluation. Researchers must be evaluated — they need to publish and then stand behind their work. WIKIPEDIA is written collaboratively by largely anonymous Internet volunteers. No academic employer will acknowledge work you contribute to an open source that is untraceable.

On the other hand, academic institutions should motivate researchers to also contribute to big non-commercial open and collaborative projects that create

public goods such as WIKIPEDIA, WorldCat and Public Library of Science. Contributions to these systems should be tracked and used to evaluate scientists, because WIKIPEDIA *does* have an impact and this impact can be measured.

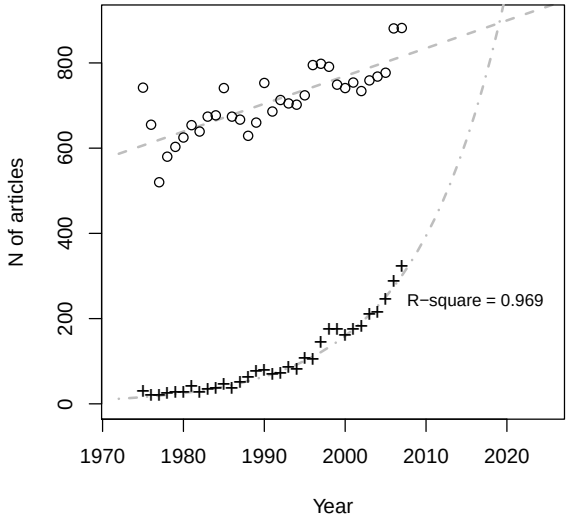


Fig. 4.5 Comparison of publication growth rates in the field of psychology for national (○) vs international (+) teams (this indicates that by 2020 there will be more trans-border author teams than national teams). This figure has been reproduced by using the script and data kindly contributed by KLIEGL and BATES (2011). In comparison with Fig. 1.3, one does not need to obtain copyright permission to generate new original visualizations and insights based on these data (hence our scientific creativity was not *'locked'*).

This document was, for example, produced using the ScribTex[36] collaborative editor. We started with a local copy. Once the book had reached 60% of the final content, we placed it on-line and then continued writing into the live version only. ScribTex contains a full TeXLive installation customized to generate PDFs even if some artwork files are missing. Other alternatives for collaborative writing of LaTeX documents are LaTeXLab and/or Subversion. Once we finished writing the book jointly, we placed a copy on A.nnotate.com to get feedback from various colleagues (read-only access). We then moved on to prepare the Camera Ready Copy of the book.

In fact, you can easily install MediaWiki on a server, then add some plug-ins that allow the export of wiki text to LaTeX or Open Office formats. There are now also efficient solutions that allow integration of programming (or any type of data processing) and document writing. For example, a package called S–weave[37] allows integration or R and LaTeX code. That way statistical analy-

[36] `http://www.scribtex.com`

[37] `http://www.stat.uni-muenchen.de/~leisch/Sweave/`

sis and production of plots and tables can be automated, so that with any new iteration the authors only need to re-run the script and it will update the content using the new data. This means that the complete research article can be put into a single code (ASCII file) and co-authors can collaborate on different aspects (statistical computing, line editing, graphics and visualization) at the same time. Consider for example an article on the internationalization of research by KLIEGL and BATES (2011). You can easily obtain the input data and R code used to generate all plots from the author's homepage and reproduce and/or extend the analysis (Fig. 4.5).

The idea of opening up both your data and the code you develop to process the data is very much in line with what Prof. Pebesma calls *"the six pillars of open research"*[38]. These are:

1. *open data, in real-time* (i.e. data on a server)
2. *open source software*
3. *open, reproducable procedures*
4. *open, web-based methods for data and processing models* (interoperability)
5. *open and explicitly quantified significance and accuracy levels of research findings*
6. *managed, open user and developer communities*.

This is probably too much to expect from every research group, but these principles are well worth considering.

Collaborative editing is becoming a standard way of producing articles, even when the co-authors are located along the same corridor. Papers produced by adopting such a process should be better because of the advantages listed above. Nevertheless, before a group jumps into collaborative writing via a web browser or by using subversion or Git, it's probably a good idea to first agree on the issues discussed from page 58 onwards.

4.10 Free and Open Source Software

> *"To build a better world we need to replace the patchwork of lucky breaks and arbitrary advantages that today determine success — the fortunate birth dates and the happy accidents of history — with a society that provides opportunities for all."*[39]

In 1983, Richard Stallman, frustrated by some poorly implemented printer drivers, launched the free software movement and two years later published the GNU Manifesto. The GNU General Public License[40] was published in 1989 and the first functional Linux operating system was released in the early

[38] Edzer Pebesma's Inaugural lecture, University of Münster, June 25, 2010. The complete talk is available at `http://www.opengeostatistics.org`.

[39] Malcom Gladwell in *"Outliers"*.

[40] `http://www.gnu.org/licenses/gpl.html`

1990s. From then on, we've seen numerous examples of Free and Open Source (FOSS[41]) software projects that have been developed in parallel with commercial versions. At the beginning of the 21st century, we can say that FOSS has a solution for most of the commercial software applications needed to run a professional business (including scientific production and publishing). Here are some examples of popular FOSS used by academia (the numbers in brackets indicate the number of unique visitors according to Google Ad planner):

- *Web-browsing*: Mozilla Firefox (210M)
- *Office work*: OpenOffice (7.3M)
- *Operating System*: Ubuntu (4.3M)
- *Database management*: MySQL (3.8M)
- *General-purpose web programming*: PHP (6.2M)
- *Web publishing and Content Management System*: Drupal (3.1M)
- *General-purpose high-level programming*: Python (1M)
- *Statistical programming*: R (320K)
- *Scientific writing and document generation*: LaTeX (120K)
- *Collaborative editing / version control system*: Subversion (110K)

Although they are available free of charge, all of these software packages are highly professional. They are maintained by active communities of developers, who are often top programmers/scientists and have an extensive publication history behind them.

[41] FOSS is a widely accepted abbreviation for software either registered under GPL or liberally licensed to grant the right of users to use, study, change, and improve its design by making its source code freely available.

For an academic or anyone working for a government agency or NGO (i.e. a public servant) it makes absolute sense to use and promote FOSS. However, experience teaches us that few people have adopted FOSS as the main software suite for their work. Why? There are several reasons. First, commercial software companies market their products more aggressively. They often spend a great deal of money on advertising in newspapers, multimedia and on the internet, but also on visits and presentations to potential clients[42]. FOSS groups, on the other hand, do not spend any money on promoting new releases.

Second, software first became a commercial business in the 1970s and 80s and many big software companies are still dominant simply because people have got used to their software and most of their documents and files are still in their formats. Even though there is now a solution that could save companies and institutions a great deal of money[43], many are slow to switch to FOSS, perhaps due to the high costs of training and data migration.

Third, FOSS requires additional skills to read the code. FOSS is, in fact, completely based on the ability of people to read the program code, i.e. understand the programming languages behind it, and then adjust and/or improve it. Human nature is such that, as with any complex system, when people cannot understand the *magic* behind the system — it frightens them.

Most people still use commercial software from Microsoft, IBM, Apple or Oracle to run their projects. Even NGOs and academic institutions mainly use commercial software. The situation is, however, changing. Today, even commercial companies are discovering the advantages of using open source software. IBM, for example, is increasingly shifting its focus to using open source software, mainly Linux for web applications. Oracle distributes and maintains the OpenOffice suite. Sun maintains Java development software. So while commercial (closed code) software is still the mainstream, there is an unstopable trend toward opening up the code. To learn more about FOSS visit WIKIPEDIA's free software portal[44].

Ironically, support received from developers of FOSS is often of higher quality then that received from many commercial software companies. As Rolf Turner puts it:

> *"In the middle of a Saturday morning (in my Time Zone!) I send out a plea for help, and in just over 20 minutes my problem is solved! I don't think you get service like that anywhere else. This R-help list is BLOODY AMAZING!"*

This book was largely produced using FOSS. The text was typeset in LaTeX, several graphs were produced using R and Open Office. Although LaTeX might seem difficult for those who are not used to looking at the code[45], you should at least try it. Maybe the professional layout of this book will convince you to consider switching to FOSS.

---

[42] Google's staff, for example, spend 70% of time on search and advertising.

[43] E.g. Microsoft Office Professional costs about 500 US$ per licence; Windows 7 professional costs about 200 US$.

[44] `http://en.wikipedia.org/wiki/Portal:Free_software`

[45] There are also user-friendly WYSIWYG editors for LaTeX such as Scientific WorkPlace.

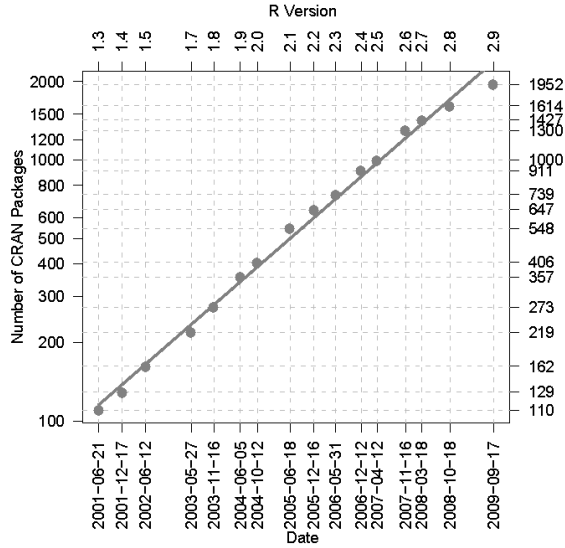


Fig. 4.6 An example of a successful academic FOSS project is the R software for statistical computing (CHAMBERS, 2008; KABACOFF, 2011). A growth in the number of contributed packages (submitted on the Comprehensive R Archive Network) has been exponential, although the core group of developers consists of only a dozen researchers/enthusiasts. Plot taken from FOX (2009, Fig.3).

We are not saying that switching to FOSS is a MUST for everyone all the time[46], but FOSS does need to be encouraged, especially if you work for government-funded projects and/or in education.

4.11 The abstract

The last stage of writing is usually to generate an abstract and choose the official title for your paper. Titles and abstracts are important because these are the parts of your article that will be read an order of magnitude more often than the article itself. So it's a good idea to spend more time on producing them. Writing abstracts requires a lot of skill, because they must be compact, complete and full of useful information. Fig. 4.7 shows Jorge Cham's view on the way people write abstracts.

[46] In this book we've also used commercial software. For example, most of the artwork was produced using Microsoft Visio and Corel Draw.

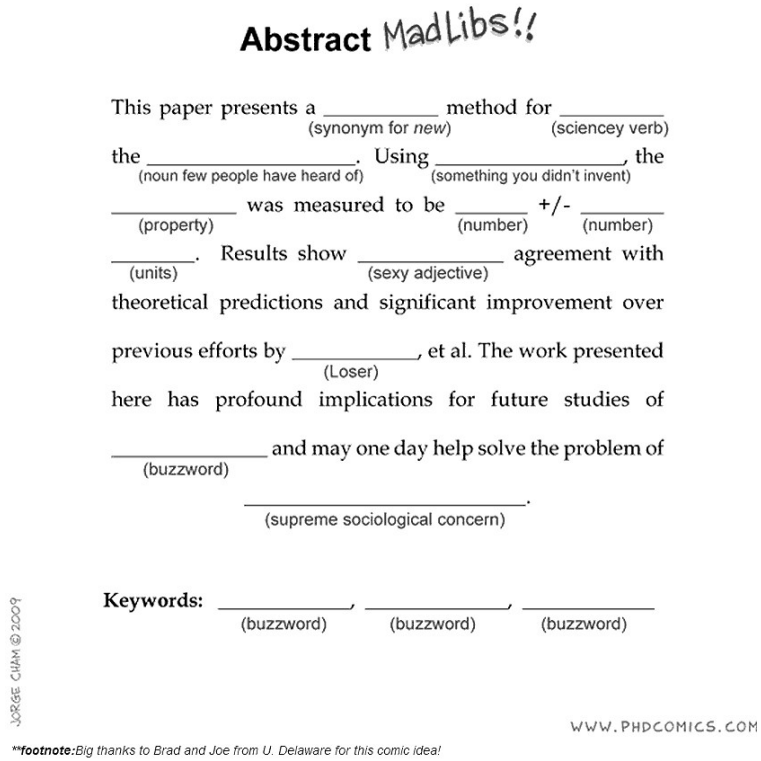


Fig. 4.7 Writing abstracts by Jorge Cham, author and creator of the PhD Comics magazine. *"Piled Higher and Deeper"* by Jorge Cham `www.phdcomics.com`.

There is about a 10 times higher probability that somebody will read the title and abstract of your paper than the paper itself. You should therefore spend about 10 times as much effort on writing (and rewriting) them.

In fact, it's a good idea to write abstracts based on a template (though not the fill-in form Cham jokingly presents). For example, you can try the six-sentence template that was used to produce the paper in the appendix:

1. *What is this paper about, what is the topic and/or key objective?*
2. *What was the experimental design?*
3. *Which methods and materials were used to analyze the data / produce the results?*
4. *What do the results show (what was discovered)?*
5. *What are the main conclusions?*
6. *What are the (broader) implications of this work?*

Abstracts are usually written in the passive form and the past tense (they describe what was done and why). In fact, the best abstract could well be written by a close colleague who is familiar with the topic and has read and understood the paper (almost like a review).

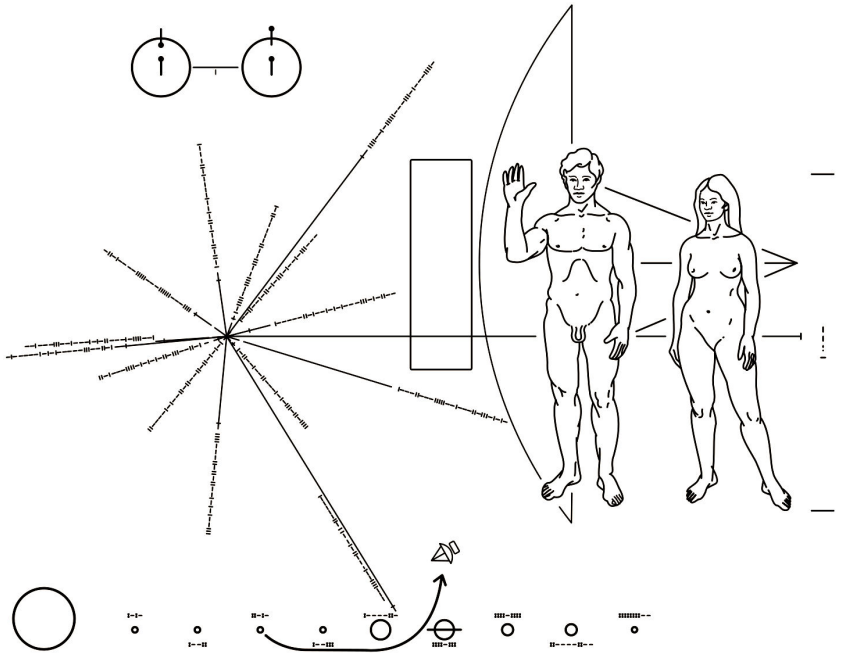


Fig. 4.8 NASA image of Pioneer 10's famed Pioneer plaque. A summary of our solar system, and of the creatures who made the Pioneer. Although it looks like abstract art, this drawing is actually packed with information.

Abstracts should be packed with technical information. See for example *"Life on planet Earth: a one-page summary"* written by Nick Trefethen[47] — this is not really an abstract, but it illustrates how to write compactly. Fig. 4.8 shows another example of an impressively compact abstract — the Pioneer plaque showing the symbolic representation of our solar system using *'universal'* measures. Too bad that researchers are not allowed to use such graphical symbols to produce summaries of their work, too.

Another thing to consider when writing abstracts is that abstracts should *'sell'* an idea. Most readers use the title and abstract to decide whether to read your paper. They are a bit like an elevator pitch. *"Elevator pitch"* reflects the idea that it should be possible to deliver the essence of an idea/proposal in the time span of an elevator ride (approximately thirty seconds to two minutes).

[47] `http://people.maths.ox.ac.uk/trefethen/`

With an abstract you get a similar opportunity: convince someone that your results are interesting using just a few lines of text.

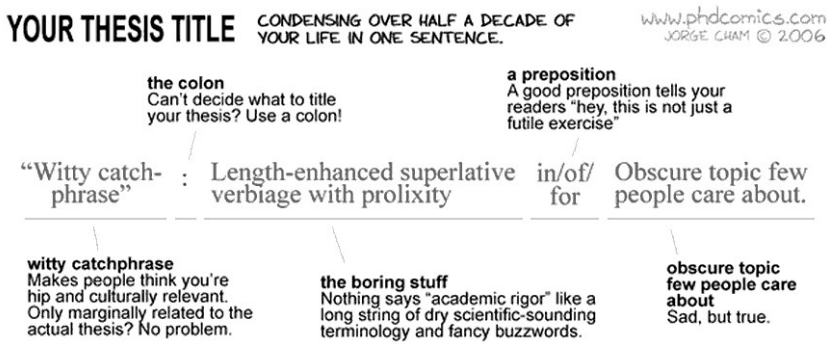
An abstract is a miniature version of the article. It should try to answer (at least) the following six questions: (1) What is this paper about? (2) Which experimental design was used? (3) Which methods and materials were used? (4) What do the results show? (5) What are the main conclusions? (6) What are the implications of this work? Abstracts should be packed with relevant information. Sentences should be short and (mainly) in the past tense.

Both abstract and title need to be accurate representations of the content of the article. If a reviewer is not able to find the information promised in the title or abstract, then their confidence will be seriously reduced. In fact, it's the ethical responsibility of authors to accurately reflect their work in the abstract and title.

4.12 The title

"The way to create art is to burn and destroy ordinary concepts and to substitute them with new truths that run down from the top of the head and out from the heart."[48]

The basic function of the title of a research paper is to accurately reflect its content and main discoveries. A title should answer the simple question *"what is this work about?"* (using the least possible number of words). Here's Jorge Cham's view of the way PhD students tend to generate titles for their theses[49]:



[48] Charles Bukowski in *"Sifting Through the Madness for the Word, the Line, the Way"* (2003).

[49] "Piled Higher and Deeper" by Jorge Cham `www.phdcomics.com`.

Indeed, many titles of research articles are a combination of catch phrases, boring expressions and terms that only a limited group of people really understand. Can we do better? Absolutely.

The best way to get inspired about generating titles is to look at inspiring examples. We will now list some titles that might give you an idea how to brainstorm a winning title. Let's start with a classic:

"An index to quantify an individual's scientific research output"[50]

This is the paper we refer to in the first part of this book, which introduces the *h*-index (named after the author). What it nicely shows is that the author takes a balanced view of its discovery: he mentions *"an"* index, and then specifies what it's for. The author could have suggested the name for this index to avoid ambiguity, but we all now know that this is the key source that describes the previously mentioned *h*-index (see page 16). We advise you to read this article not only because of its title, but also because it's a good example of a well-written compact paper that made history.

Here's another example of a good title:

"Testing Hubbert"[51]

This paper shows extensive results from testing the Hubbert oil-reserves calculation model (a bell-shaped curve in which the area under the curve is equal to estimates of the total amount of oil available) using global oil production data. It won the Best Student Paper Award Competition at the 26th North American Conference of the International Association for Energy Economics, not only because of its intriguing title, but also because it is high-quality scientific work. What's so good about it? Several things:

1. It's intriguing — it promises an interesting read.
2. It's short and catchy.
3. In fact, it's so short and catchy that people notice it. It's as if the author has invented a new style for making titles. Originality is always inspiring.

The author could have easily used a longer, more explanatory title such as:

"Statistical evaluation of the Hubbert model for prediction of the total remaining amount of oil in the world"

But the author doesn't try to explain what *"Hubbert"* means. Everyone who reads that journal is already familiar with Hubbert's curve, so there's no need to explain it. Even someone who's not an expert in the field could figure out that it is about Hubbert's curve by quickly Googling the term. It's also clear that word *"testing"* refers to statistically matching actual measurements with the model. By removing redundancy the author creates something that is catchy, while being completely accurate about the topic of the paper.

---

[50] by J.E. Hirsch; published in 2005 in the PNAS journal.

[51] by Adam R. Brandt; published in 2006 in the journal Energy and Policy.

Trimmed titles are not always better. For example, BEEL and GIPP (2009) discovered that Google Scholar's ranking algorithm puts high weighting on words in the title. A longer title that includes all keywords might improve the chances that people will be able to find your work, so consider using the main keywords; then remove all redundancy, personal (and/or professional) bias, and ambiguous buzzwords from the title.

Probably the best way to generate a title of a research article is to find a balance between using the key terms / discoveries and length. Titles can be also improved by removing redundancy and bias — emotionally loaded words, overselling statements and ambiguous buzzwords.

Here's a longer example:

"Wealth and happiness across the world: material prosperity predicts life evaluation, whereas psychosocial prosperity predicts positive feeling"[52]

This paper explores the connection between happiness, income, and psychological needs. Although the title seems to be rather long, it has an original element: it puts the most important conclusion in the subtitle and hence efficiently conveys the main message. We almost don't need to read the abstract! It's an original research paper that reports novel results, so it helps to see these results in the title.

Do buzzwords work? — Buzzwords are new fashionable terms that are used in the media. Buzzwords typically have a negative connotation among scientists because they indicate that some originally technical term is now used *"pretentiously and inappropriately by individuals with little understanding of its actual meaning."* Some typical examples are: cyberspace, sustainability, nanotechnology, cloud computing, and E-learning. Intuitively, a buzzword can increase the visibility of your paper — buzzwords are catchy. But you should also be critical. First, you need to think of the buzzword as a dynamic thing: what will happen to it in five to ten years' time? Will it still be as popular? Some quite recent buzzwords now seem ridiculous.

Here's a somewhat different example, this time the title of a book:

"Does God Play Dice? The New Mathematics of Chaos"[53]

In this example the author uses a famous quote by Albert Einstein to intrigue his audience. A title without this quote would have been as accurate but, because it's a book on popular science, an intriguing title will attract a wider audience. In research articles, it's often also effective to place a big research question directly in the title. Just make sure you provide an unambiguous answer to that question.

Another example, this time from The Economist:

[52] by Diener *et al.*; published in 2010 in the Journal of Personality and Social Psychology.

[53] A book by Stewart, I. published by Blackwell in 1989.

"Organising the web: The science of science"[54]

This is an article about developments in the field of web-crawling and automated categorization. Probably too general a title, because science and scientific systems cannot be simplified to web-crawling. If you look at this article you will notice that it only presents some initial results from a small group at Princeton University. This tells us more about the authors[55] of the article than about the topic.

Based on these examples, we can roughly classify journal article titles into six groups:

- *Topic in the title* — The title here is basically the field of study. The advantage of such a title is that it can be relatively short; the disadvantage is that the title is not dynamic and is probably more general than the content of the paper.
- *Topic : subtitle* — A somewhat better version of the topic-type title is a title with a sub-title that refers to the main discovery/key claim.
- *Claim in the title* — The title is the key claim. See the example by Diener *et al.* on the previous page.
- *Research question in the title* — A title can also be formulated as the key research question, finishing with a question mark. This type of title is somewhat unconventional — not acceptable in all fields, hence risky.
- *Sub-field in the title or subtitle* — An ambitious way to generate titles is to literally invent new sub-fields. This might not be well received in every discipline, but it is a good idea to attach a subtitle (to an ambitious title) specifying the focus of your study.
- *Free title* — A title that does not belong to any of the above categories can be classified as an *"free title"*. Some researcher like to demonstrate that complex thinking and a sense of humor go well together. Is it appropriate to demonstrate a sense of humor in the title of a research article? Usually not, but it's a challenge you could consider. Journals typically do not welcome controversial titles, but if you have really good results you might just get away with it.

Nancy Ackles[56] suggests three general strategies for naming a paper:

I The title should be a noun phrase, not a sentence or a noun clause.
II The title should indicate the topic of the paper.
III The title should be unambiguous and use only standard English word forms.

[54] `http://www.economist.com/node/18618025`

[55] Interestingly enough, The Economists does not publish the names of authors with their articles (hopefully a model no scientific journal will follow), so we don't even know who stands behind this article.

[56] `http://depts.washington.edu/eslinfo/Lab/titles.html`

There is also the issue of redundancy, which is a common mistake many authors make. Most editors agree that you should avoid using titles that contain expressions such as *"Some new results"*, *"A novel method for"*, *"An investigation of"*, *"Effects of"* etc. If you want to trim your title, this is where you should start. And always check the heading style in the journal you have targeted to make sure your heading matches that style.

It's the author's responsibility to use a title that closely reflects what is presented in the manuscript. You don't want to either undersell or oversell (too general topic in the title/too many buzzwords). If you oversell, someone might read your article and get disappointed. Unlike in show business[57], in science there *is* such a thing as bad publicity.

The title should exactly cover the content of the article and/or express its main claim. Nothing more, nothing less.

4.13 Final tips

"The only end of writing is to enable the readers better to enjoy life, or better to endure it."[58]

The first draft of your article will probably just be *"an incoherent stack of notes you've written to yourself"* (CLARK, 2003). But this should not worry you because most high-quality papers go through ten or more editing iterations. In fact, for Zissinger, the essence of writing is rewriting. As CREEDY (2008) also puts it: *"Perhaps the most important rule of writing is that the first draft is not the final draft but is simply the start of a long process of revision."*

Even if you are just putting the pieces together, you should check whether your first draft contains the following main elements (after Ed Hull):

- *The 'big picture'* — the background to your research and why it's important.
- *The purpose statement* — what you set out to show.
- *How you went about answering the research question* — what you did.
- *The answer to the research question* — what you concluded.
- *The answer to the* so what? *question* — why should the reader care?
- *The consequences for future research* — which could change the original 'big picture'.

For more detailed tips 'n tricks, see our *"Rules of thumb for writing research articles"* in the appendix. If you follow these guidelines closely, you should be

[57] People in show business often say that there's NO such thing as bad publicity: all publicity is good.

[58] Samuel Johnson in his review of Soame Jenyns' *"Free Enquiry into the Nature of the Origin of Good and Evil"*.

able to produce a first draft of your article, which you can then revise and polish to *'perfection'*.

When you re-read your paper, you may still be unhappy with it. It might have many novel elements, but if you feel that something is missing — excitement, enlightenment, sparkle… call it what you will, such papers will not make much of an impact in science. Go back to your title, introduction and discussion and see if you can strengthen the story you are telling in your article. The points that you are trying to make need to be clear and strong. You should not be afraid of contributing new thoughts. And don't forget that scientific writing is both a science and an art. As Samuel Johnson put it: *"What is written without effort is generally read without pleasure."*

If you find line editing difficult, you could consider getting professional help. There is a thriving article editing business (KAPLAN, 2010). There are now numerous professional companies[59] that charge a reasonable price (in the range 300–500 US$ per article) for reading your paper and giving you comments and suggestions.

An experienced writer will usually take at least three to five iterations until he/she produces a version that is ready for submission. If you and your co-authors are not longer sure what to leave out and what to extend, this is a sign that you should now send the paper to the journal. Time is also a factor — science rewards only those who publish first. It's not a good idea to keep going round in circles with your co-authors. Or as Blaise Pascal once said: *"thinking too little about things or thinking too much both make us obstinate and fanatical."*

[59] Some international scientific editing companies: the Nature Publishing Group Language Editing in London, MacMillan Scientific communications, Carter's Strategic Solutions, and American Journal Experts.

Chapter 5
Submitting your article

Prior to submitting your article, we strongly recommend seriously investigating the journal (looking at the editorial board, previous issues, list of authors and their academic level) in which you intend to publish your work. There are still many things that you need to consider to avoid your paper ending up in the editor's waste bin.

5.1 Find the right journal

> *"We are at a turning point in our history. Universal access is the goal. And the opportunity of leading a different life, based on this is… thrilling. It could be one of the things humankind would be most proud of."*[1]

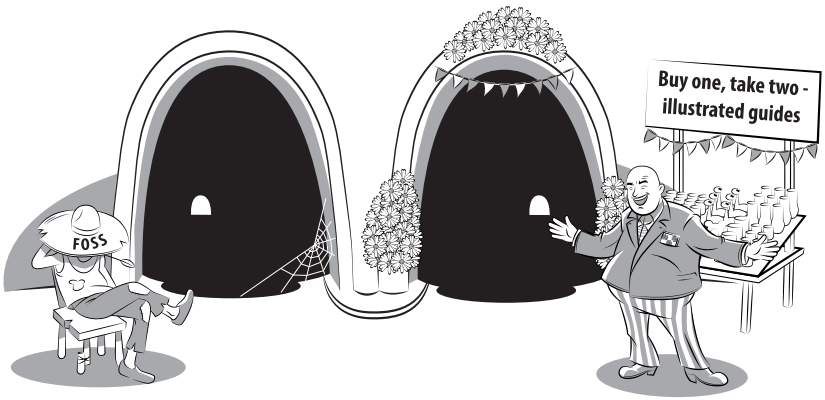
Once you have prepared a one-page concept paper, you should already have a good idea about where to publish your full manuscript. Your paper needs to closely match the scope and format of the journal; otherwise, even if it is a very promising paper, it may well be rejected (right message in the wrong place). If you have a list of possible journals, now is the time to think about which is the best candidate. The ultimate tip is: always go first for the highest quality journal that may be interested in your topic. If this journal rejects the paper, you can always turn to the second on your list. But be realistic. Do not send articles of limited interest to top journals, otherwise you will waste people's time (including your own). The key question should be: who do I want to read my paper and which journals are those people scanning?

BABOR *et al.* (2008) suggest the following key steps when selecting a journal:

- *Decide first whether the article is suitable for an international audience, either in a generic, disciplinary or specialized journal*

[1] Brewster Kahle cited by Lawrence Lessig in *Free Culture*.

- *Explore the compatibility between your article and the journal's culture (read the journal's mission statement)*
- *Make sure the journal is currently interested in the type of article you have written*
- *Gauge your exposure by reviewing the journal's circulation and abstracting service*
- *Consider, but don't get fooled by, impact factors*
- *Take into account time to publication and other practical matters.*



Scimago Labs provides a service called *"SCImago Journal and Country rank"*[2], which allows comparison of journals using a number of bibliometric indices — citations per year, *h*-index, international collaboration, etc — which makes it relatively easy to find out which journals in the field should be next on your list.

Do not transfer your copyright to a commercial publisher too easily, because they will exploit it to eternity. There are plenty of reasonably priced Open Access publishers such as BioMed Central, PLoS, Copernicus, and MDPI, which are often indexed in SCI and have increasing visibility.

The following checklist can help you judge both the quality of a journal and the editorial process:

- Check if the journal is indexed (by Thomson Reuters in the SCI or CC database). You can do this at any time from Thomson Reuters's Master jour-

[2] `http://www.scimagojr.com`

nal list[3] and/or via the SCImago website. If the journal is listed by Thomson Reuters, this will increase the chances that your article will be seriously considered. You can also check the impact of the journal or even the impact of the country[4] where you intend to publish.
- Check if the article (or at least its abstract) will be available on-line in PDF format. This will increase the chances that it will be widely accessed.
- Check if your article will receive a unique identifier, such as a **Digital Object Identifier**[5], which is something like an ISBN for books. This will make it easier to locate.
- Try to find out whether the journal provides English language and graphics editing. Journals usually do not do this, but you can often find out from colleagues. This will increase the chances that the article will be of high publication quality.
- Check if the journal has an on-line editorial system. This will help ensure that your article is not kept on hold for a long period of time.
- Try to find out if the journal offers a double-blind review and/or publishes review findings. This will help ensure that you receive an unbiased review.
- Check the Journal's (publisher's) copyright and self-archiving policy[6] — is the journal *'green'*? Do they permit public access to author's pre-prints and/or post-prints?

Note that each journal has its own review system, preferred style and specific jargon. You should always make sure your paper meets the requirements of your target journal.

5.2 Do the work of the reviewers yourself

> *"Hard work is a prison sentence only if does not have meaning. Once it does, it becomes the kind of thing that makes you grab your wife around the waist and dance a jig."*[7]

Another useful pre-submission tip is to try to do the work of the reviewers yourself (put yourself in reviewer's place) or give the paper to your colleagues and ask them to review it. Once you have created some psychological distance from the draft paper, you can read your work more critically. Reviewers usually complain about similar things — either they are frustrated trying to understand your argument, they are not sufficiently convinced, or they are disappointed that you are not acknowledging their work.

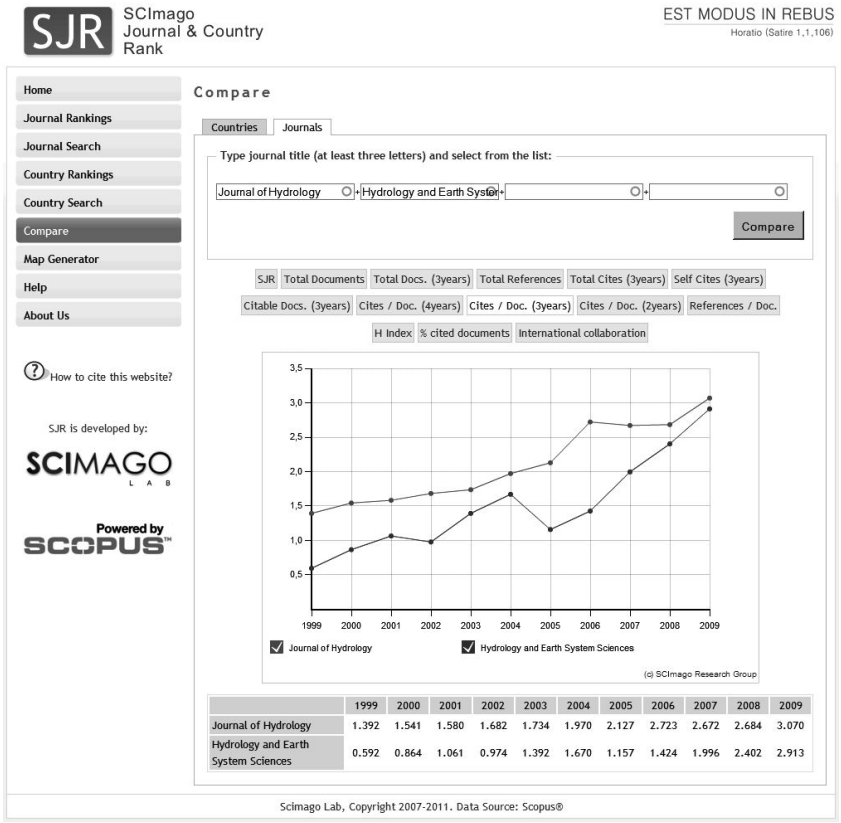
[3] `http://science.thomsonreuters.com/mjl/`

[4] For example, the country with the highest number of citations per article in the last 10 years is Switzerland. It is also the country with the highest number of citations per researcher. See also `http://sciencegateway.org/rank/`.

[5] See `http://doi.org` for more details.

[6] `http://www.sherpa.ac.uk/romeo/`

[7] Malcom Gladwell in *"Outliers"*.



	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009
Journal of Hydrology	1.392	1.541	1.580	1.682	1.734	1.970	2.127	2.723	2.672	2.684	3.070
Hydrology and Earth System Sciences	0.592	0.864	1.061	0.974	1.392	1.670	1.157	1.424	1.996	2.402	2.913

Fig. 5.1 Comparison of citations per year for articles published in one STM (Hydrology) and one OA (HESS) journal. Although many commercial publishers seem to offer more service and quality (at least that's what they advertising says), low-budget publishers can often do an equally good job.

The quality of a paper is determined by the quality of the *spices* used to prepare it. For example, the quality of a new methodological paper can be determined by evaluating whether:

✓ the input data sets are of high quality (50%[8])
✓ cross-evaluation of results is available (25%)
✓ at least one alternative approach has been considered (10%)
✓ a variety of validation indices have been used (10%)
✓ a clear and honest discussion of limitations is provided (5%).

This means that you can make a stronger argument if you:

[8] This number is an empirical estimate.

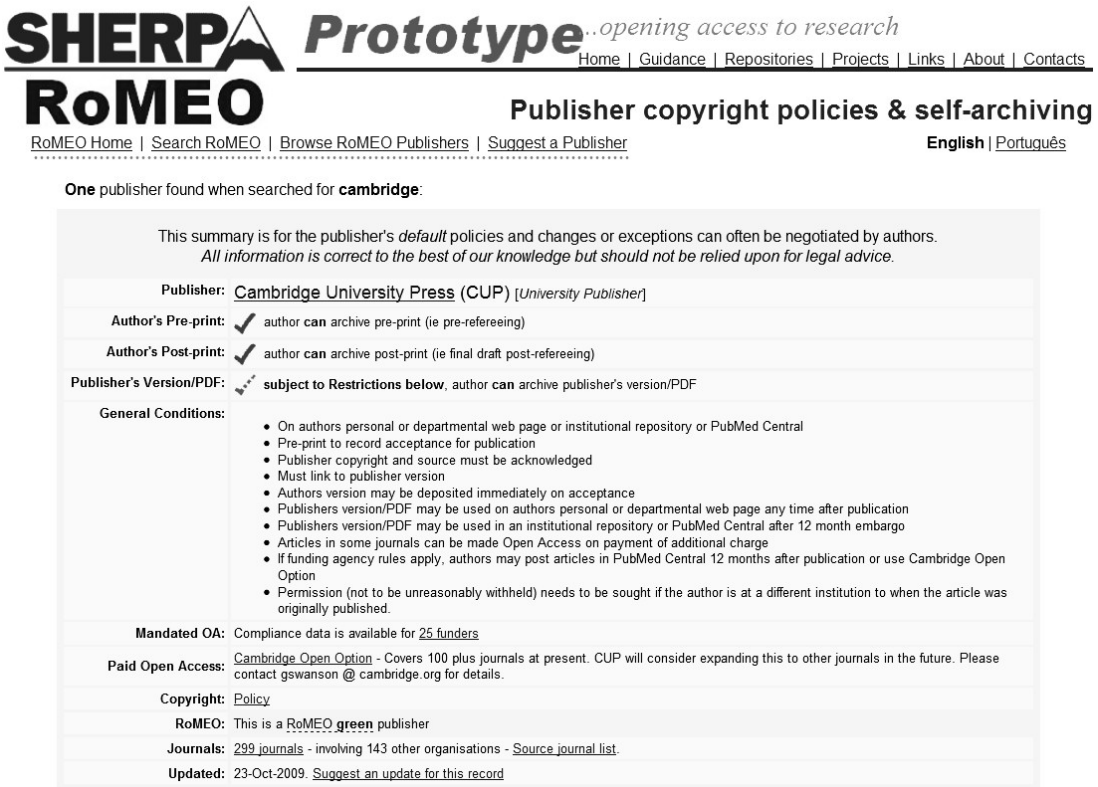


Fig. 5.2 SHERPA (Securing a Hybrid Environment for Research Preservation and Access) provides an overview of publisher's policies. Worth checking before you select a journal.

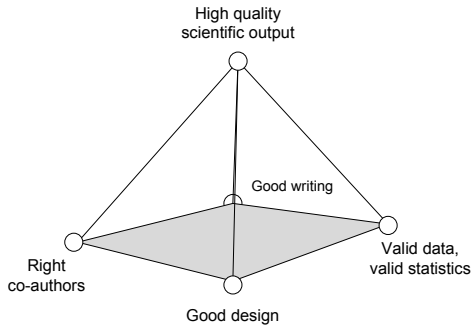


Fig. 5.3 What makes a good paper? A combination of the right co-authors, good design, good data and good writing. See also the 10 rules for getting published on page 63.

Consider alternative solutions
Compare your method with alternative methods using rigorous criteria. Try to describe both the strong and weak aspects of your method in relation to alternative techniques. Recall the rules of science on page 5: report your discoveries in an unbiased way; you should even try to challenge your own initial ideas. Big problems are seldom solved with a single study.

Test it under different conditions
Evaluate the performance of your method using several case studies or experiments. How does the method work under different conditions, both global and local? How universal are your discoveries/conclusions?

Emphasize possible applications/implications
If you give concrete guidance to readers on how to apply your methodology, this will certainly increase its impact.

Identify yourself with a broader research group
Think of a research group as an international company in which you have shares. You need to support the work of your colleagues and find your identity (your *niche*) in that company. This means that you need to be self-critical and acknowledge the fact that other colleagues might find better solutions. Modest opinions and statements are usually more accurate, but avoid too much hedging — making weaker claims than those your paper justifies.

You may need to return to your data and even do some extra data collection. Although you might not feel like doing this, think about how you would react if you were to receive a negative review (rejection or serious revision needed). You might lose six or more months waiting for advice that you can anticipate now.

Make a multi-level description of your method/results — go from the simplest to a more specific case and then on to a general case. A simple case study

- You have carefully and systematically **studied the work of colleges** in the field and referred to it
- You have **critically evaluated** your methods/results versus state-of-the-art findings
- You have tested the performance of your technique using **multiple case studies / experiments**
- You **make significant claims**, but these are backed up by strong arguments
- You have considered what the **implications of this work** are and how these ideas/discoveries can be used to **solve real-life problems**
- You have **adjusted your style and *jargon*** to the target audience

- You do not actually know what has been published by others on this topic
- You are sure that your technique is optimal and that there is no better alternative
- You do not know if you will get the same results under various conditions
- You're not prepared to make significant claims or you are making them without sufficient proofs
- You do not actually know how your technique can be applied to solve real-life problems and what the implications of your findings are
- You have not previously tried to communicate your ideas/results to your research group

Fig. 5.4 Some Do's and Don'ts to consider before submitting your article.

(that shows real numbers) will help readers understand the technique, while the multiple case studies will convince them that it works under a variety of conditions. The important thing is that readers can zoom in and out, depending on their interest.

Submitted research papers should not be draft documents but double and triple-checked manuscripts that are almost ready to go to print. You might think that some minor errors can be fixed later on, but for the reviewer just one fatal flaw (!) can excuse them from checking all of the subsequent details (SMITH, 1990). SMITH (1990) is, in fact, a must-read for anyone who wants to understand how reviewers look at the papers.

5.3 Make a full information package

"Some books seem to be written not to instruct, but rather to show that the author knows something."[9]

Reviewers get you published, but it's readers who get you cited. Prepare additional promotional material such as websites, posters, and brochures that can help people understand your work or convince them that the results are correct. In business, this is called *post-production* and it's particularly well developed in show business — Hollywood film companies, for example, spend enormous amounts of money promoting their new products.

Graphical elements can play a key role here because scientific information is often communicated more easily through figures. Sometimes it can help your paper if you include extra diagrams — even sketches or photographs. Whatever helps you convey your message and draws attention to it is good. Your final product might include:

- The core paper itself (typically about 15 pages of text).
- Technical notes or supplementary materials in which data collection, analysis and interpretation are evaluated in detail.
- A website where all supplementary materials/datasets can be found and accessed. A website is also useful for providing multimedia such as animations and videos. On the other hand, it is not a good idea to put your whole manuscript on the web before it has been indexed. *'Gray'* publications are soon forgotten and there is a risk that someone else might take credit for your work.
- Promotional material such as brochures and posters, in which your key ideas/discoveries are presented in a popular way. An excellent tool is a pubcast in which you talk to camera about your work and then link this clip to your paper[10].

According to Prof. Edzer Pebesma (University of Münster) scientists can increase their impact if they attach datasets and publish software that goes with their articles — *"People love to use software and/or datasets that come with a reference — it makes them feel they do science. And to us, it's the number of citations that get us somewhere."*

There is now a general notion to publish the whole research so that anyone can reproduce and advance it. This is becoming increasingly important for scientific fields such as climate change and biodiversity assessment (KLEINER, 2011).

---

[9] Johann Wolfgang von Goethe 1749–1832 in *"Maximen und Reflexionen II"*. This quote reflects the main difference between writing at university (for your professor) and for a real-world audience.

[10] `http://www.scivee.tv/pubcasts`

5.4 Pre-submission checklists

"Progress in research is largely achieved by making a series of small steps, rather than taking giant leaps."[11]

Even though your paper is starting to look complete and serious, you should not rush into submitting it to a journal. You can use the following checklist[12] to take one last look at the paper:

- *Do I still like the beginning?* Ideas evolve as you write — sometimes it's best to write the Introduction last.
- *Can I write more clearly, concisely and simply?* Avoid complex jargon, and replace pompous with plain language. Look at the text from the reader's point of view. Does it need to be so complex?
- *Do I make sense?* Check for contradictions, ambiguity and poor logic.
- *Do my numbers add up?* Check all figures and statistics carefully, or get them checked. If two numbers do not add up, reviewers start to rapidly lose confidence even if these are accidental typos.
- *Do my sentences* hang *together?* Write smooth, dynamic paragraphs (not all the same length).
- *Do my verbs pull their weight?* Use strong verbs, keep them close to their subjects and make unnecessary passive formulations active.
- *Do I need every modifier?* Delete non-essential adjectives and adverbs / keep them close to nouns and verbs.
- *Have I got rhythm?* Listen to the sound of your writing. It should be easy to read out loud. If it is not — re-organize, compress and improve connections in the text.
- *Am I playing in tune?* Listen to the tone. Is it in harmony with your subject and your audience?
- *Can I trim?* If in doubt, cut it out. Say things once, not twice or three times.
- *Have I made my case?* Step back and assess the power of your arguments. Ask a colleague's opinion.
- *How's my grammar and spelling?* Don't forget to run your spelling checker. But be skeptical about your software' grammar checker — it's not foolproof.

Finally, you should check your contact details and affiliation. For affiliation, many recommend that you should avoid using a third level classification. For example, the most common affiliation is (1) university/institute + (2) research group. Adding more levels simply confuses people because most of the academic systems in the world are based on a *legal entity* (company owner) + *research group* (project / educational programme). Adding more than two classification levels can also lead to inaccuracies because most indexing services have only two levels in their database.

---

[11] John Creedy in *Research without tears*.

[12] Adapted from O'CONNOR and WOODFORD (1975).

5.5 The review process and response to reviews

"When you meet someone better than yourself, turn your thoughts to becoming his equal. When you meet someone not as good as you are, look within and examine your own self."[13]

Before or during the review process, you should also consider some of the following tips, which might be crucial for the success of your article:

- *Check out your editors* — Investigate the people who will decide about your work: editors, established researchers and other potential reviewers. Check all the little things that annoy them and try to deal with them. If you want to ask them something about your paper, do it in a concise and concrete way. Editors are extremely busy people.
- *Be ready for unfortunate developments* — Be prepared to receive biased, harsh or even offensive reviews. It's possible that your paper will be perceived as weak because it misses the most important point (recall the rules of science on page 5) or does not have a real audience. Otherwise, if you have received a poor review, consider trying a higher quality journal. Recall rule of science No. 1: there is no authority in science, only the power of arguments.
- *Work with your heart, write with your head* — You need a lot of passion to do research and write research publications, but being emotional when communicating with editors, reviewers and other colleagues is something you should always avoid.
- *Don't get discouraged and give up* — Fighting for a publication is normal. Once you submit your article, this is only the beginning of the game. As Tara K. Harper puts it: *"Live. Write. Finish your work. Be willing to accept critique. Be willing to improve"*[14]
- *Be honest about your work* — Create some distance from it and critically evaluate it. Then, try to improve or even re-design the paper. If you can't do this, then just be honest and mention the limitations of your findings. Recall the rules of science on page 5: even the most pessimistic truth is better than an illusion.

Reviews can be roughly classified into five categories: (1) *minor corrections*, (2) *needs rewriting* and line editing but the data/argument is probably fine, (3) *moderate revision* (reorganization) required and the authors need to provide more evidence and enrich the paper with better graphics and tables (come back in few months), (4) *major reworking / redesign needed* — even new data — but the editor likes the idea (come back in half a year), (5) *reject* — this paper is not suitable for this journal and/or is not novel and/or focuses on a topic of minor interest and/or is flawed.

If your paper gets rejected or the editors asks for major revisions, there are three possible scenarios:

[13] Confucius in *The Analects* (cca. 475 BCE — 221 BCE).

[14] `http://www.tarakharper.com/faq_pub.htm`

You have sent it to the 'wrong' journal
As already mentioned, the choice of journal can be crucial. If the topic of your paper does not fit the scope of the journal, of course they will reject it. Similarly, the quality of your paper should match the quality of the journal. PhD students cannot expect to get a positive response if they send their first paper to Nature. Nature rejects >90% of submitted papers by default.

You have received an unjustified lousy review
This does not happen too often, but it may happen because of, for example, bias or simply because a reviewer had a bad day. You can consider writing to the editor to ask for a second review. If the editor disagrees with you without sufficient justification, then you should give up on this journal and send the paper to a higher quality journal. Sometimes, if the review you received is lousy, but the paper has been accepted for publication anyway, you should still consider sending it to a higher quality journal.

The reviewers are correct and clear
The best thing to do in a situation like this is to give up on the current version and completely change course. Often when we work on the same topic for an extensive period of time, we can no longer see it objectively. Some people become obsessed with their work. It's human nature to be biased about our own beliefs and concepts. Many researchers simply cannot give up parts of a project they have been working on for a long time. In such situations, consult some senior colleagues and be honest with yourself. Researchers should be very flexible about considering a change of course or even a complete change of topic.

How can you tell if the reviews are lousy? First, the minimum requirement is that the reviewer reads the paper from the beginning to end. If it is obvious from the comments that he/she did not do this, then it will be difficult to proceed with any discussion. Studies have shown that reviewers who do not spend more than three hours on a review do not increase the quality of the papers involved (EVANS *et al.*, 1993). Second, a lousy review contains criticism without evidence and argumentation. Especially if you notice emotionally loaded words, you should consider writing back to the editor and request more justification or even third opinion. The same way authors have to provide logical reasoning, reviewers also need to follow the same principles.

Anyone can get their paper rejected. There's nothing dramatic about getting research papers rejected. It can happen to even the most experienced writers. Especially if you are new to publishing you should not take it too hard. Rejection doesn't necessarily mean that you are not suited to an academic career. Rejection is usually the result of a wrong strategy: wrong journal, wrong message and/or wrong audience. If this can be of any comfort, publishing scientific work is not as stressful as publishing a novel — less than 1% of all manuscripts submitted for publication in the USA each year actually get published[15].

[15] According to `http://www.tarakharper.com/faq_pub.htm`.

Even if you get rejected, a good referee report is immensely valuable and the reviewers don't charge for it (SMITH, 1990; TURABIAN, 2007). So, even if reviews are poor or superficial, you should appreciate the work others have done and try to make the best use of it. GARDNER (2006) has promoted the idea of five basic cognitive abilities (intelligences) — a disciplinary mind, a synthesizing mind, a creative mind, a respectful mind, and an ethical mind — you should have all five, especially a respectful mind.

5.6 PhD study: a leap in the dark

"Sail away from the safe harbor. Catch the trade winds in your sails. Explore. Dream. Discover."[16]

Previous sections contain tips on how to improve scientific writing from the macro to the micro level. In this final section we focus on PhD students. *"Starting a PhD thesis is typically a leap in the dark. This naturally leads to anxieties"* (CREEDY, 2008). In many western countries, supervisors expect a PhD student to produce a thesis as a *compilation of (at least) three research articles* on a specific problem or research topic HARTLEY (2000); HARTLEY and BETTS (2009). If an author publishes three articles in international journals, he/she has demonstrated an ability to design, practice and publish science, and is therefore worthy of receiving a doctorate. People may even start to identify you with some key terms — you've started to establish your *niche* in the world of science.

A typical academic career looks like this (see also Fig. 5.5):

1. First, three articles published in international peer-reviewed journals
2. PhD promotion
3. Post-doctoral projects (the next 10 important publications)
4. Organize research conferences or collaborative projects (e.g. editing books)
5. Permanent position or research tenure
6. Popularize developments in your field in the media
7. Initiate (get funding for) your own projects and educational programs, and hire new PhD students and post-docs
8. Become a senior researcher / professor.

Initiation i.e. getting your first publication in a peer-reviewed journal is not a straightforward process and success is not guaranteed. Even if you were an excellent student at university, there's no guarantee that you will easily get published. There are some major misconceptions that need to be clarified:

- Your supervisors will not take responsibility if your paper is rejected — the responsibility will be entirely yours. Once you become a senior researcher or a professor you will discover how stressful it would be to take responsibility

[16] Quote by Mark Twain.

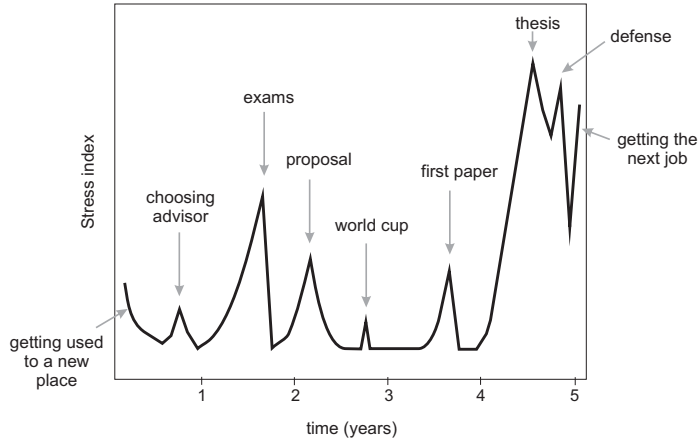


Fig. 5.5 The Graduate School stress curve by Edwin P. Gerber (from a talk on how to survive grad school).

for everything your students produce (there is simply not enough time to go through all the material in detail).

- Likewise, many supervisors do not have time to read every text you write in detail and filter it according to the steps described in section 1.6 (page 24). Do not think that if your supervisor has no specific comments, your paper is okay and ready to publish. It probably isn't. You should read section 1.6, and go through the checklists listed there to minimize the chances of rejection.
- Journal editors are not part of your PhD committee and they are not interested in any arrangements you may have with your supervisors. The papers you intend to submit must be stand-alone documents that target the readers of the journal. After you publish a paper, you can make some minor modifications and adjust it so that it links more closely to your PhD thesis. Such adjusted papers can be published as *"Based on… "*.
- Journals will not fit in with your work tempo or project plans. Your paper can easily be delayed (even unintentionally), so make sure you submit papers to journals in the first half of your PhD study or at least 18 months before the planned defense.
- PhD students are *at the bottom of the food chain* in the academic kingdom. Appreciate your supervisors and always be very professional (organized) when sending documents and undertaking work-related tasks.

CREEDY (2008, p.67–68) thinks that the worst things a PhD student can do when communicating with supervisors are (paraphrased):

- *Be unwilling to make suggested changes and/or carry out extensive revisions to drafts*
- *Immediately claim that the supervisor is wrong about something*

- *Send piles of computer output and expect the supervisor to sieve through all the detail*
- *Send scrappy or highly provisional pieces of work and then expect extensive feedback*
- *Ignore or treat lightly suggestions by supervisors*
- *Ask to borrow books.*

In conclusion, if you're well organized, becoming a PhD is not as hard as it may seem. BLOOMFIELD and EL-FAKAHANY (2008, p.329) make a number of suggestions on how to get organized:

- *Set up a workspace with good working conditions*
- *Follow an outline*, to ensure the most logical arrangement of topics
- *Don't worry about getting all the wording right in the first draft*
- *Start with writing things that are easiest* (methods and materials)
- *Print out drafts and reread them after a break* (critically evaluate yourself)
- *Make back-ups!*

WEINBERG (2003) suggested four golden rules for PhD students at the start of their careers:

1. *Don't be too anxious about a lack of background knowledge (no one knows everything, and you don't need to!).*
2. *Go for the messy (unsolved problems... mysteries) — that's where the action is.*
3. *Forgive yourself for wasting time. Unsuccessful experiments and dead-ends are quite normal.*
4. *Learn something about the history of science. Take a broad-brush approach*[17].

You should have the guts to take risks. Learn how to swim in rough waters. Experimenting is OK if you are a PhD student, even if you find you need to flush some of your hard work and start again.

When you are presenting your work at an international conference you might get a chance to impress people with your presentation skills and slides. Scientific publication has to achieve the same result without all the bells and whistles. As we have emphasized several times in this book, the best way to convince an audience of your credibility is to produce a scientific publication of the highest quality. Researchers are now also experimenting with different ways to enhance the impact of their work by using, for example `pubcasts` — short visual explanations of your research[18]. Maybe this will help you express your ideas clearly and convince your colleagues?

---

[17] Consider, for example, reading Carl Sagan's *"Cosmos"* and/or Bill Bryson's *"A Short History of Nearly Everything"*.

[18] See for example `http://www.scivee.tv` — a channel for drawing attention to your research results through video presentations linked to your paper. All you need is a WebCam.

Finally, the key to success with a PhD study probably lies in getting professional help from your co-authors, i.e. your supervisors. If we were to put all of the ingredients needed for a successful PhD in order of importance, it would look something like this (the percentages in brackets indicate estimated weighting):

- At least two supervisors who have the right advice at the right time and you can sit in their office and talk with them whenever necessary (50%)
- Top laboratory and IT equipment, including software (20%)
- Access to digital and hard copy libraries (15%)
- Specialized training in academic writing, research skills, advanced statistical analysis and document preparation (10%)
- Participation in international conferences and workshops (5%).

About 50% of a successful PhD thesis is down to having good supervisors, who are usually co-authors of your articles (CREEDY, 2008). And this typically does not cost much. If you prepare a solid draft or concept paper, then there's a better chance that you will be able to attract influential researchers and get them on board. Researchers are best rewarded by publications. If your supervisors or external colleagues recognize the potential in your paper, they will want to get involved. So, as with love, money can't buy you a PhD. Or to quote the Naked Chef: *"I am not a doctor. I'm a chef. I do not have expensive equipment or medicine. I use information, education."*[19] You too can prepare an excellent meal with a low budget.

[19] `http://www.ted.com/talks/jamie_oliver.html`

Recommended reading

On Free and Open Source Software manuals:

★ Ayers, Ph., Matthews, C., Yates, B., 2008. **How Wikipedia Works**. No Starch Press, 536 p.
★ Chambers, J.M. 2008. **Software for data analysis: programming with R**. Springer, 231 p.
★ Kabacoff, R.I., 2011. **R in Action: Data analysis and graphics with R**. Manning publications, 470 p.
★ Mittelbach, F., Goossens, M., Braams, J., Carlisle, D., and Rowley, C. 2004. **The LaTeXCompanion**, 2nd edition, Addison-Wesley Professional, 1120 p.
★ Gay, J. (edt) 2002. **Free Software, Free Society: Selected Essays of Richard M. Stallman**. GNU project, 226 p.

On graphic design and scientific visualization:

★ Goossens, M., Mittelbach, F., Rahtz, S., Roegel, R., and Voss, H. 2007. **The LaTeXGraphics Companion**, 2nd edition, Addison-Wesley Professional, 976 p.
★ Tufte, E.R., 1992. **The Visual Display of Quantitative Information**. 2nd Edition, Graphics Press, 197 p.
★ Sarkar, D., 2008. **Lattice: Multivariate Data Visualization with R**. Springer, Use R series, New York.

On writing (and finishing) a PhD thesis:

★ Creedy, J. 2008. **Research without tears: from the first ideas to published output**. Edward Elgar Pub., 118 p.
★ Dupre, L. 1998. **BUGS in Writing**, Revised Edition: A Guide to Debugging Your Prose. 2nd ed., Addison-Wesley Professional, 704 p.
★ Turabian, K.L. 2007. **A Manual for Writers of Research Papers, Theses, and Dissertations**. 7th ed., University of Chicago Press, Chicago, p. 436.

Appendix

Try to pick a catchy title!

Rules of thumb for writing research articles

Tomislav Hengl[a] & Michael Gould[b]

[a] ISRIC — World Soil Information, PO Box 363, 6700 AJ Wageningen
[b] Michael Gould Associates BV, Apeldoornseweg 21, NL-6814 BG Arnhem

Abstract—**This paper lists 'rules of thumb' for writing research articles (RA) and getting them published. These were discussed during a scientific writing course organized for ITC PhD students by Cressie Communication Services. Important aspects of the macro and substructure of a paper were explored in group discussions. The (meso)structure and functions of different sections of RAs are described. The results of previous investigations and interviews among journal editors were used to identify what makes a good RA. It was concluded that clear, logical, coherent, balanced and well-structured writing gets papers published and read. Some important rules of the thumb were: *Adjust your writing to the audience and purpose*, *Avoid redundancy and unnecessary explanations* and *Write like you speak and then revise*. These rules can help inexperienced writers present their work in a more effective way.**

An abstract should be short but give the overall idea: what was done, what was found and what the main conclusions are

Key words: research article, writing, rules of thumb, structure,

Introduction

MOVE 1: Introduce the topic and emphasize why it is important!

A scientific or research article or paper is a technical (or essayistic?) document that describes a significant experimental, theoretical or observational extension of current knowledge, or advances the practical application of known principles (O'CONNOR and WOODFORD, 1975). It is important to emphasize the fact that a research article (further referred as RA) should report on research findings that are not only sound (valid) and previously unpublished (original), but also add some new understanding, observation, proofs, i.e. potentially important information (GORDON *et al.*, 1983).

Unlike a novel, newspaper article or an essay, an RA has a fixed, predefined structure and style, which is by international consensus known as *Introduction-Methods-Results-Discussion* or IMRaD. However, an RA is not only a technically rigid document, but also a subjective intellectual product that unavoidably reflects personal opinions and beliefs. Therefore, it requires skill both in structuring and formulating concepts and findings. These skills are acquired through experience, but can also be taught.

MOVE 2: Relate to current knowledge:

There are many books on general guidelines for scientists wishing to write RAs and survive the review process (TRELEASE, 1969; DAY and GASTEL, 2006; GERMANO, 2008). These days, many scientific societies publish quite detailed style manuals to help both authors and publishers; see for example the CBE's Style Manual (1994) or the ACA-CSA-SSSA Manual (2011). In this paper, the conventions for writing an RA are described in the terms of the macro, meso and micro-elements of the paper.

"What's been done" and "What need's to be done?"

Various authors have investigated the principles of creating a good title, writing a good abstract or introduction (MCPHEE, 2001; SWALES and FEAK, 2004). Some go to the level of the micro-structure of the RA (sentences), providing guidelines for structure and style (GOPEN and SWAN, 1990; TURK and KIRKMAN, 1989; KIRKMAN, 2004).

Introduce the GAP

However, writing an RA is a bit like climbing a *monkey-puzzle tree*, especially if you are a non-native English speaker (further referred to as L2). What makes a good paper and can we offer guidelines to young researchers?

Objective

With this motivation, we tried to formulate some rules of thumb for writing (and publishing) RAs. These were gathered during discussions in the course *"Scientific writing for non-native English speakers"*, but they also reflect our personal experience of scientific writing. The idea behind this paper was to summarize the main conclusions from these discussions in an easy-to-read format.

MOVE 3: Introduce your work Give the purpose and main objective

Note that we do not focus on correctness. Rather, we try to show how authors would be more/less likely to write in a particular way in a specific context. The need for unambiguous clear rules, rather than fuzzy preferences, is probably culturally determined (HOFSTEDE *et al.*, 2005). Fortunately, the structure of scientific writing is well-defined and we can also indicate what is effective style (in terms of readability).

Methods and materials

The Scientific writing course, organized annually for ITC PhD students, was held between March 8 and April 26 in 2002. There were nine students, who followed five full-day classes. There was enough time to do homework — mainly

rewriting assignments in between classes. The classes were organized in such a way that the participants worked in groups or individually and discussed the most important issues, first among themselves and then with the whole group.

Describe the experimental set-up

The following topics were discussed in detail (in chronological order): the standard structure or elements of an RA, macro, meso and micro levels of an RA, general problems with readability and communication, the functions and content of the Introduction, Methods, Results and Discussion sections, writing successful abstracts and tips for submitting an RA for publication in a journal. The participants were from eight countries (L2) and four continents, which provided a good basis for discussing cultural-academic differences (PRINCE *et al.*, 1999).

Focus of the study

The material and facilities were organized by Ian Cressie 2011, while most of the classes were led by Michael Gould, documentation consultant and advisory editor. The participants generated some graphs and flow diagrams manually (Fig. 5.6), which we then modified and transferred to a manuscript form.

Explain the methodology

Fig. 5.6 Photo from the Scientific writing class at ITC. Discussion about the *"Discussion"* section.

An RA is like a cook-book! Be specific and provide all the necessary details

The basic concept of the course is that the students should learn from real examples and their own mistakes. In most cases, participants were analyzing and correcting each other's work. In other cases, participants were making comments on examples prepared by Ian Cressie. A typical exercise was, for example: a short RA is given to students who have to write an abstract, respecting the appropriate conventions.

Establish an author's 'voice'

Most of the rules mentioned in this article were agreed by the majority of participants. We have also used the results of previous investigations and inquiries by journal editors to support general conclusions. Neverthe-

less, some of the statements and principles reflect personal views and opinions and should not be confused with the cited literature. The tips given here apply mainly to application-based sciences and RAs intended for publication in such journals.

Results

RA structure and style

Provide a summary of results

A RA was first divided into a number of sections (futher referred to as RAS) and elements (RAE). The participants agreed that the main article sections that are required in any modern journal are, in this order: Title, Authors, Abstract, Introduction (I), Methodology (M), Results (R), Discussion (D) and Conclusions and References. These are the core of an RA. Additional listed RAS's were: Author-paper documentation, Keywords, Acknowledgements, Abbreviations and Appendices. The RAEs listed were: tables, figures (plus graphs, maps, diagrams, and sketches), equations, citations and footnotes and comments. The RAEs can come in different places in the RA. However, tables and figures are more usual in the Results section and equations and citations in the Methods section and Introduction. All of these RAS's and RAEs have their function and required style and should form a coherent whole. The functions of the main RAS's are described in Table 5.1.

Compare results

The participants agreed that some RAs, even with good data and interesting results, will be rejected if the style and format of the paper are not tailored to the needs of a particular audience. This confirms the results of GOSDEN (1992, Fig.1) who asked over 100 journal editors what they thought were the most important issues for non-English authors who want to get published. These were, in order of priority:

1. Clear, logical linking of sentences
2. Coherent development of the topic (old before new information)
3. Use of grammatically correct sentences
4. An ability to make effective claims at the right level
5. Clear organization of sections of a paper, and
6. Placing their work in a wider context (especially relevant for authors in developing countries).

The misplacement of old and new information is not just a problem for non-native speakers, it is also the No. 1 problem in American professional writing (GOPEN and SWAN, 1990; GERMANO, 2008). The participants analyzed a flawed paper by an unknown author and decided, after some discussion, that they would reject the publication.

Table 5.1 Research Article Sections (RAS), main functions, preferred style and related rules of thumb.

RAS	Main functions	Preferred style	Rules of thumb
Title	• Expresses the content and main discoveries • Attracts the reader's attention	• Short and simple words (7-10) • Purposive (targets a specific audience)	• Avoid complex grammar; make it catchy! • Avoid redundancy ("*An investigation of...* ", "*The analysis of...* ", "*Effect of...* ", "*Influence of...* ", "*New method...*")
Abstract	• Reflects the main 'story' of the RA • Calls attention but avoids extra explanations	• Past (perfect) tense and passive voice • Short concise sentences • No citations, tables, equations etc.	• Avoid introducing the topic • Explain: what was done, what was found and what are the main conclusions • Provide summary 'numbers'
Introduction	• Introduces the topic and defines the terminology • Relates to existing research • Establishes the focus of the paper and research objectives	• Present tense for referring to established knowledge and past tense for literature review • Extensive overview of literature	• Use state-of-the-art references • Follow logical moves • Define your terminology to avoid confusion • Establish an author's voice
Methodology	• Provides enough detail for competent researchers to repeat the experiment • Who, what, when, where, how and why?	• Past tense • Correct and internationally recognized style and format (units, variables, materials etc)	• Mention everything you did that influenced the results; don't cover your traces ("*some data was ignored*") • If a technique is widely known, refer to it by its name (don't re-explain it) • Use the simple(st) examples to explain complex methodology

Table 5.2 Research Article Sections (RAS), main functions, preferred style and related rules of thumb (continued).

RAS	Main functions	Preferred style	Rules of thumb
Results	• Provides summary results in tables and graphs • Compares different 'treatments' • Provides quantified proofs (statistical tests)	• Past tense • Use tables and graphs and other illustrations	• Present summary data directly related to your research question and not all research results • Call attention to the most significant findings • Make a clear distinction between your work and that of others
Conclusions and Discussion	• Clearly answers research questions/objectives • Explains discrepancies and unexpected findings • States importance of discoveries and future implications	• Simple past and/or present tense (past tense when related to results) • Allows scientific speculation (if relevant)	• Do not recapitulate results but make statements • Make strong statements where they are justified (avoid "*It may be concluded...* " style) • Do not hide unexpected results — they may be the most important ones
References	• Provide a list of related literature and sources of information • Support the ideas in the paper	• Depends on journal, but authors/editors, year and title must be included	• Always cite the most accessible references • Cite primary source rather than review papers

RA sub-structure

Focus: put more focus on what should be emphasized

The participants also discovered that all RAS's can be separated into subsections using clear signposts, which can improve the way the argument is built up in an RA. The subsections we identified were: research topic and definitions, research objectives (questions), methodological techniques, experimental set-up, context of the study (e.g. study area), main discoveries (analyzed data), answers to research questions, explanation of the conclusions and further research and implications. The main RAS's are listed in a flow chart, showing the main relationships between the different sections (Fig. 5.7). Fig. 5.8 shows the substructure of the Introduction and Discussion — the most important RAS's.

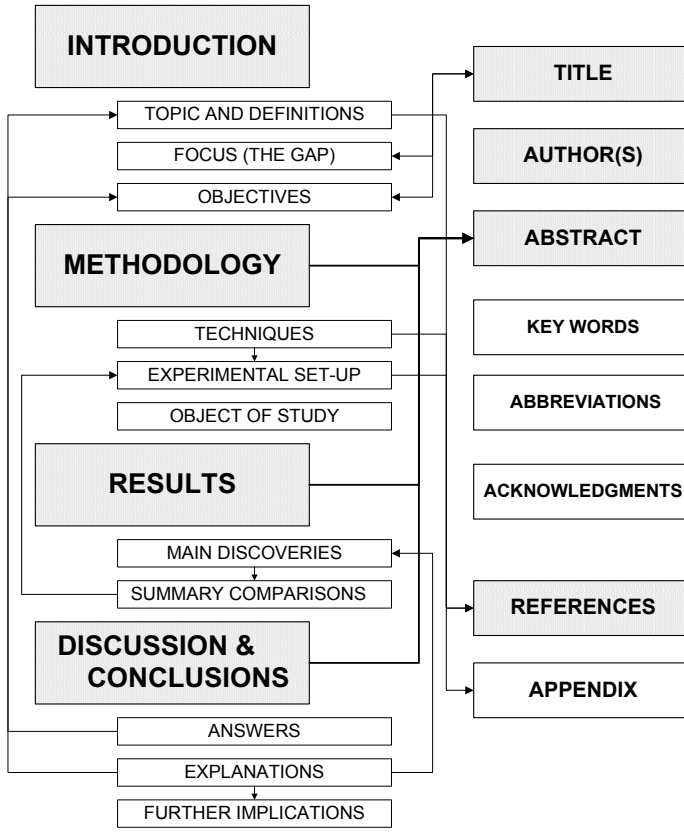


Fig. 5.7 Flow diagram: research article sections (shaded) and subsections, and their main relations.

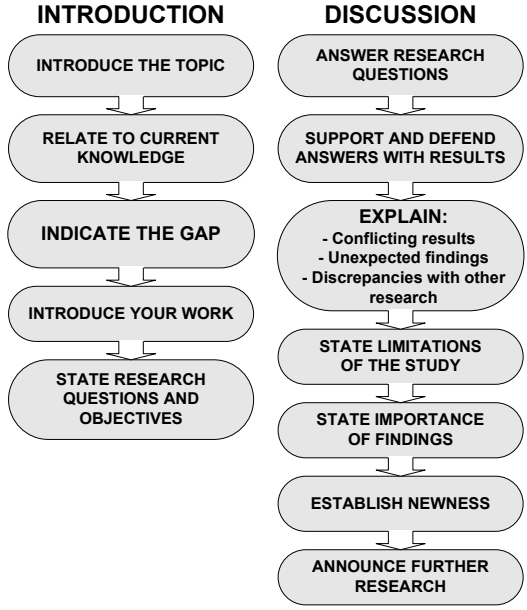


Fig. 5.8 Flow diagram: logical framework for the Introduction and Discussion, as agreed by most of the participants.

Discussion and Conclusions

Answer the research question

Provide summary conclusions

What is the purpose of an RA and what makes it a good one, and who decides that it is a good RA? Are there rules for easier writing? If the main function of an RA is to transfer new knowledge on a research topic, then a good paper is one that is clear, coherent, focused, well argued and uses language that does not have any ambiguity. However, it is not only the message that is important. The RA must have a well-defined structure and serve as a kind of cook book, so that others can reproduce and repeat the experiments described in it. There are some rules that can make the writing and publishing of RAs easier. Here, we summarize some which should always be kept in mind by an inexperienced researcher (Table 5.3). We put all of these together to make a list of some 40 logical steps, which can be found at the end of this article.

Unexpected findings

Although it was assumed in the past that *'thicker'* articles with a wider range of vocabulary are preferable, most editors (and readers) prefer simple, clear and coherent writing (KISS — Keep It Short and Simple), rather than a fancy or complex,

Table 5.3 Selected golden rules for easier publishing.

Rule	*Explanation*
Take a reader's view	Write for your audience not for yourself.
Tell a story	Direct your RA but keep a clear focus in the paper and present only results that relate to it.
Be yourself	Write like you speak and then revise and polish.
Make it simple	Use simple(st) examples to explain complex methodology.
Make it concrete	Use concrete words and strong verbs, avoid noun clusters (more than three words), abstract and ambiguous words.
Make it short	Avoid redundancy, repetition and over-explanation of familiar techniques and terminology.
Take responsibility	Make a clear distinction between your work and that of others.
Make strong statements	*"We concluded… "* instead of *"It may be concluded… "*
Emphasize	Learn to use little words in a big way.
Be self-critical	Consider the uncertainty of conclusions and their implications and acknowledge the work of others.
Never stop editing	The key to writing well is extensive editing.

pseudo-scientific style. FUNKHOUSER and MACCOBY (1971) showed that information gain is especially enhanced by the *use of examples*, i.e. it helps a lot to use parallels from everyday life, historical references, etc. Some sections, such as the Introduction and Discussion, must intrigue readers, and capture their interest. For example, an interesting title can catch readers' attention and will be easily remembered (e.g.: T.Y. Li and J. Yorke named their famous paper on chaos: "*The period three means chaos*"). Some sections simply require more skill and are more important.

It is estimated that of all the published journal RAs in the world, less than 5% are read in detail. However, more than 50% of abstracts are read and so the quality of an abstract is much more important. Therefore, the abstract should present the 'story' of the RA in miniature and should make sense stand-alone.

The sub-structure of an Introduction was first described by SWALES (1981) with his "*four moves*". These later on become three, the so-called CaRS (Create-A-Research-Space) model , which are: establish a research *territory*, establish a research *niche*, and then occupy the niche (SWALES and FEAK, 2004). The participants in the course concluded that especially the meso-structure of the Introduction and Discussion should follow a logical flow of *'moves'*, similar to playing chess (Fig. 5.7 & 5.8).

The more structured and precise a paper is, the greater the chance that it will get published. Each of the RA elements has to fulfil its function in order to achieve this goal.

Establish newness

The importance of following a logical structure is nicely illustrated by GOPEN and SWAN (1990): "*People need signposts to understand what you're communicating. First establish the context based on what they know. Then move towards the new facts you want to convey. Beginning with exciting new information and ending with something we already know leaves us disappointed and spoils the flow*."

However, this is not the whole story. An RA must target a specific audience/journal, has to be novel and of high interest. Finally, one thing should be uppermost in researchers' minds: a good article is not only an article that has been published in a top journal — it is the reaction it generates that makes the difference.

Explain discrepancies

Therefore, a good article is not just one that is well written. A good article is one that is read and cited. In some cases, even a good paper will be rejected. Unfortunately, sometimes the reasons for this can be subjective (perhaps in around a third of all cases). Editors are often biased, they prefer one or another approach, academic level, gender. . . nation. These problems and issues such as fraud, plagiarism and ethics are not discussed here, but they certainly are important.

Searching, inputting and formatting references has been much improved lately with the help of so-called *"information management tools"* (web applications, on-line libraries, reference management tools, etc.). In addition, the role of companies involved in sorting and filtering, such as the Thompson's Web of Knowledge, will become more important.

Further research and implications

In the future, we can expect more structured guidelines for writing an RA (perhaps even templates?). The RA will also probably support multimedia (animations, sound recordings), which will improve communication between authors and readers/users. These innovations will inevitably require some new rules of thumb.

We would like to thank Ian Cressie for arranging the courses at ITC, which are of great importance to L2 PhD students. We also thank former ITC PhD student Jose L.C. Santos and Dr. David G. Rossiter for reading the text and making their valuable suggestions.

Put it all together: writing an RA in 40 STEPS

MAKE DRAFT

STEP 1 Draft a working **title**
STEP 2 Introduce the **topic** and define terminology
STEP 3 Emphasize why the is topic important
STEP 4 Relate to **current knowledge**: what's already been done
STEP 5 Indicate the **gap**: what needs to be done?
STEP 6 Pose research **questions**
STEP 7 State your overall purpose and **objectives**
STEP 8 List methodological **steps**
STEP 9 Explain the **theory** behind the methodology used
STEP 10 Describe the **experimental set-up**
STEP 11 Describe the technical details
STEP 12 Provide summary **results**
STEP 13 Compare different results
STEP 14 Focus on main **discoveries**
STEP 15 Answer research questions (and draw clear **conclusions**)
STEP 16 Support and defend **answers**
STEP 17 Explain conflicting results, unexpected findings and **discrepancies** with other research
STEP 18 State the **limitations** of the study
STEP 19 State the **importance** of your findings
STEP 20 Establish **newness**
STEP 21 Announce **further research**

REVISE

STEP 22 ABSTRACT: what was done, what was found and what are the main conclusions
STEP 23 Is the title clear and does it reflect the content and main findings?
STEP 24 Are key terms clear and familiar?
STEP 25 Are the objectives clear and relevant to the audience?
STEP 26 Are all variables, techniques and materials listed, explained and linked to existing knowledge - are the results reproducible?
STEP 27 Are all results and comparisons relevant to the stated objectives?
STEP 28 Do some statements and findings repeat in the text, tables or figures?
STEP 29 Do the main conclusions reflect the questions posed?
STEP 30 Will the main findings be acceptable to the scientific community?
STEP 31 Is the text coherent, clear and focused on a specific problem/topic?
STEP 32 Is the abstract readable standalone (does it reflects the main story)?

POLISH

STEP 33 Are tenses used appropriately (including the active and passive voice)?
STEP 34 Are all equations mathematically correct and explained in the text?
STEP 35 Are all abbreviations explained?
STEP 36 Reconsider using words such as "very", "better", "may", "appears", "more", "convinced", "perfect", and "impression" in the text.
STEP 37 Are all abbreviations, measurement units, variables and techniques internationally recognized (IS)?
STEP 38 Are all figures/tables relevant and of good quality?
STEP 39 Are all figures, tables and equations referred to in the text?
STEP 40 Are all references relevant, up to date and accessible?

Fig. 5.9 The 40 steps you need to follow to write an RA.

References

Cited sources:

ADLER, R., EWING, J., and TAYLOR, P. (2009). Citation Statistics. *Statistical Science* **24**, 1–14.

ADRIENKO, N. and ADRIENKO, G. (2005). *Exploratory Analysis of Spatial and Temporal Data, A Systematic Approach*. Springer.

ASA-CSA-SSSA (2011). *Publications Handbook and Style Manual*. ASA-CSA-SSSA, Madison (WI).

ASCHERON, C. and KICKUTH, A. (2004). *Make Your Mark in Science: Creativity, Presenting, Publishing, and Patents, A Guide for Young Scientists*. Wiley-Interscience.

BABOR, T. F., STENIUS, K., SAVVA, S., and O'REILLY, J., editors (2008). *Publishing addiction science: a guide for the perplexed*. International Society of Addiction Journal Editors, Essex, UK, 2nd edition.

BAILLARGEON, N. and SCHMIDT, A. (2008). *A short course in intellectual self-defense*. Seven Stories Press.

BAR-ILAN, J. (2008). Which h-index? — A comparison of WoS, Scopus and Google Scholar, Scientometrics. *Scientometrics* **74**, 257–271.

BAUER, K. and BAKKALBASI, N. (2005). An Examination of Citation Counts in a New Scholarly Communication Environment. *D-Lib Magazine* **11**, 1–4.

BEEL, J. and GIPP, B. (2009). Google Scholar's Ranking Algorithm: The Impact of Articles' Age (An Empirical Study). In *Proceedings of the 2009 Sixth International Conference on Information Technology: New Generations*, pages 160–164, Washington, DC, USA. IEEE Computer Society.

BLOOMFIELD, V. and EL-FAKAHANY, E. (2008). *The Chicago Guide to Your Career in Science: A Toolkit for Students and Postdocs (Chicago Guides to Academic Life)*. University Of Chicago Press.

BOLLEN, J., VAN DE SOMPEL, H., HAGBERG, A., BETTENCOURT, L., CHUTE, R., RODRIGUEZ, M. A., and BALAKIREVA, L. (2009). Clickstream Data Yields High-Resolution Maps of Science. *PLoS ONE* **4**, e4803.

BOURNE, P. E. (2005). Ten Simple Rules for Getting Published. *PLoS Comput Biol* **1**, e57.

BOYACK, K. (2009). Using detailed maps of science to identify potential collaborations. *Scientometrics* **79**, 27–44.

CHAMBERS, J. (2008). *Software for data analysis: programming with R*. Statistics and computing. Springer.

CHOMSKY, N., BARSAMIAN, D., and BAILLARGEON, N. (2002). *Entretiens avec Chomsky*. Écosociété.

CLARK, L. (2003). *The Writing Coach*. Thomson Delmar Learning, Clifton park, NY.

COLLINS, J. (2005). The future of academic publishing: What is open access? *Journal of the American College of Radiology* **2**, 321–326.

COMMITTEE, C. S. M. (1994). *Scientific style and format: the CBE manual for authors, editors, and publishers*. C B E STYLE MANUAL. Cambridge University Press.

CORBYN, Z. (2010). To be the best, cite the best. *Nature* .

CREEDY, J. (2008). *Research without tears: from the first ideas to published output*. Edward Elgar Pub.

CRESSIE, I. (2011). *Writing & Editing Course, Lecture notes for ITC students*. ASA-CSA-SSSA, Enschede.

CRUMP, T. (2002). *A brief history of Science*. Costanble & Robinson Ltd, London.

DAY, R. and GASTEL, B. (2006). *How to write and publish a scientific paper*. Cambridge University Press.

DUPRÉ, L. (1998). *BUGS in writing: a guide to debugging your prose*. Addison-Wesley.

ERREN, T. C., CULLEN, P., ERREN, M., and BOURNE, P. E. (2007). Ten Simple Rules for Doing Your Best Research, According to Hamming. *PLoS Comput Biol* **3**, e213.

EVANS, A., MCNUTT, R., FLETCHER, S., and FLETCHER, R. (1993). The characteristics of peer reviewers who produce good-quality reviews. *Journal of General Internal Medicine* **8**, 422–428.

EVANS, J. A. and REIMER, J. (2009). Open Access and Global Participation in Science. *Science* **323**, 1025.

FEYNMAN, R. (1965). *The Character of Physical Law*. Cox Wyman Ltd., London.

FOX, J. (2009). Aspects of the Social Organization and Trajectory of the R Project. *The R Journal* **1**, 5–13.

FUNKHOUSER, G. R. and MACCOBY, N. (1971). Communicating Specialized Science Information to a Lay Audience. *Journal of Communication* **21**, 58–71.

GARDNER, H. (2006). *Five minds for the future*. Leadership for the common good. Harvard Business School Press.

GARFIELD, E. (1979). *Citation Indexing — Its Theory and Application in Science, Technology, and Humanities*. John Wiley & Sons, Inc. NY.

GARFIELD, E. (1990). The most-cited papers of all time: SCI 1945–1988. *Current Contents* **7**, 3–14.

GERMANO, W. (2008). *Getting it published: a guide for scholars and anyone else serious about serious books*. Chicago guides to writing, editing, and publishing. University of Chicago Press.

GILES, J. (2005). Start your engines. *Nature* **438**, 554–555.

GLENN, J., GORDON, T., and FLORESCU, E. (2009). *State of the Future*. The Millenium Project.

GOPEN, G. D. (2004). *Expectations: Teaching Writing from a Reader's Perspective*. Pearson Education Ltd.

GOPEN, G. D. and SWAN, J. A. (1990). The Science of Scientific Writing. *American Scientist* **78**, 550–558.

GORDON, M., MARTLEW, R., and CENTRE, P. C. R. (1983). *Running a refereeing system*. Aids to scholarly communication. Primary Communications Research Centre, University of Leicester.

GOSDEN, H. (1992). Research Writing and NNSs: From the Editors. *Journal of Second Language Writing* **1**, 123–39.

HARTEMINK, A. (2002). *Publishing in soil science — Historical developments and current trends*. IUSS, Vienna.

HARTLEY, J. (2000). Nineteen ways to have a viva. *PsyPAG Quarterly Newsletter* **June**, 35.

HARTLEY, J. and BETTS, L. (2009). Publishing before the thesis: 58 postgraduate views. *Higher Education Review* **41**, 29–44.

HARZING, A.-W. (2010). *The Publish or Perish Book*. Tarma Software Research Pty Ltd, Melbourne, Australia.

HARZING, A.-W. and VAN DER WAL, R. (2008). Google Scholar: the democratization of citation analysis? *Ethics in Science and Environmental Politics* **8**, 62–71.

HARZING, A.-W. and VAN DER WAL, R. (2009). A Google Scholar h-index for journals: An alternative metric to measure journal impact in economics and business. *Journal of the American Society for Information Science and Technology* **60**, 41–46.

HIRSCH, J. E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 16569–16572.

HOFSTEDE, G., HOFSTEDE, G., and HOFSTEDE, G. (2005). *Cultures and organizations: software of the mind*. Successful strategist series. McGraw-Hill.

INFORMATION AND COMMUNICATION UNIT (2006). *Study on the economic and technical evolution of the scientific publication markets in Europe*. European Commission Directorate-General for Research, Brussels.

JASPER, J. (2002). Why So Many Academics are Lousy Writers. *Chronicle Careers* **3-26**.

KABACOFF, R. (2011). *R in Action: Data analysis and graphics with R*. Manning publications.

KAPLAN, K. (2010). Publishing: A helping hand. *Nature* **468**, 721–723.

KIRKMAN, A. (2004). *Good Style: Writing for Science and Technology*. Routledge Study Guides Series. Routledge, Abingdon.

KIRKMAN, J. (2001). Third person, past tense, passive voice for scientific writing. Who says? *European Science Editing* **27**, 4–5.

KLEINER, K. (2011). Data on demand. *Nature Climate Change* **1**, 10–12.

KLIEGL, R. and BATES, D. (2011). International collaboration in psychology is on the rise. *Scientometrics* **87**, 149–158.

KURZWEIL, R. (2005). *The Singularity Is Near*. Penguin Group.

LARIVIÈRE, V. and GINGRAS, Y. (2010). The impact factor's Matthew Effect: A natural experiment in bibliometrics. *Journal of the American Society for Information Science and Technology* **61**, 424–427.

LATOUR, B. (1987). *Science in Action — How to Follow Scientists and Engineers through Society*. Harvard University Press, New York.

MCPHEE, L. (2001). Teaching the Research Article Introduction. In *First EATAW Conference: Teaching Academic Writing Across Europe*, Groningen.

MEHO, L. I. and YANG, K. (2007). A New Era in Citation and Bibliometric Analyses: Web of Science, Scopus, and Google Scholar. *Journal of the American Society for Information Science and Technology* **58**, 1–21.

MONTERO, I. and LEÓN, O. G. (2007). A guide for naming research studies in Psychology1. *International Journal of Clinical and Health Psychology* **7**, 847–862.

MOORE, K. (2005). Publish and flourish. *CSA News* **6**, 11–14.

NATURE EDITORIAL (2005). Not-so-deep impact. *Nature* **435**, 1003–1004.

NEWMAN, J. (2000). Twenty of the greatest blunders in Science in the last twenty years. *Discover Magazine* **21**.

NEWMARK, P. (2003). Peer review and the rewards of open access. *Nature* **422**, 661–661.

O'CONNOR, M. and WOODFORD, F. (1975). *Writing scientific papers in English*. Associated Scientific Publishers.

O'DOWD, A. (2011). Peer review system needs thorough evaluation, MPs hear. *BMJ* **342**, d3046.

PRINCE, A., BLICBLAU, A. S., and SOESATYO, B. (1999). *Implicit Actions and Explicit Outcomes: Cultural-Academic Interactions in Writing Journal Research Articles*. Swinburne University of Technology.

RAYMOND, J. (1993). I-Dropping and Androgyny: The Authorial *I* in Scholarly Writing. *College Composition and Communication* **44**, 478–483.

REES, M. (2004). *Our final century: will civilisation survive the twenty-first century?* Arrow.

REGAZZI, J. (2004). The Shifting Sands of Open Access Publishing, a Publisher's View. *Serials Review* **30**, 275–280.

ROEDIGER, H. L. (2006). The h index in Science: A New Measure of Scholary Contribution. *The Academic Observer* **19**, 1–4.

ROSSITER, D. (2009). *Research Concepts & Skills*, volume 1-3. University of Twente, ITC, Enschede.

SARKAR, D. (2008). *Lattice: Multivariate Data Visualization with R*. Springer, New York.

SCHULTZ, D. M. (2010). Are three heads better than two? How the number of reviewers and editor behavior affect the rejection rate. *Scientometrics* , 277–292.

SEGLEN, P. (1997). Why the impact factor of journals should not be used for evaluating research. Education and debate. *British Medical Journal* , 314–497.

SIEGFRIED, T. (2010). Odds Are, It's Wrong. *Science News* **177**, 25–26.

SMITH, A. J. (1990). The Task of the Referee. *Computer* **23**, 65–71.

SOKAL, A. and BRICMONT, J. (1998). *Fashionable Nonsense: Postmodern Intellectuals' Abuse of Science*. Picador, New York.

SOKAL, A. D. and BRICMONT, J. (1997). *Impostures intellectuelles*. O. Jacob, Paris.

SWALES, J. (1981). *Aspects of article introductions*. Aston ESP Research Reports. Univers. of Aston, Language Studies Unit.

SWALES, J. and FEAK, C. (2004). *Academic writing for graduate students: essential tasks and skills*. Michigan series in English for academic & professional purposes. University of Michigan Press.

THE PLOS MEDICINE EDITORS (2006). The Impact Factor Game. *PLoS Med* **3**, e291.

THE WELLCOME TRUST (2004). *Costs and business models in scientific research publishing*. SQW Limited, London, UK.

THEUS, M. and URBANEK, S. (2008). *Interactive graphics for data analysis: principles and examples*. Series in computer science and data analysis. CRC Press.

TRELEASE, S. (1969). *How to write scientific and technical papers*. The M.I.T. paperback series. 115. M.I.T. Press.

TUFTE, E. (1992). *The Visual Display of Quantitative Information*. Graphics Press, 2nd edition.

TURABIAN, K. (2007). *A manual for writers of research papers, theses, and dissertations: Chicago style for students and researchers*. Chicago guides to writing, editing, and publishing. University of Chicago Press.

TURK, C. and KIRKMAN, A. (1989). *Effective writing: improving scientific, technical, and business communication*. Routledge Study Guides Series. E. & F.N. Spon.

WAGER, E. (2007). Authors, Ghosts, Damned Lies, and Statisticians. *PLoS Med* **4**, e34.

WAINER, H. (1984). How to Display Data Badly. *The American Statistician* **38**, 137–147.

WEBSTER, G. D., JONASON, P. K., and OROZCO SCHEMBER, T. (2009). Hot Topics and Popular Papers in Evolutionary Psychology: Analyses of Title Words and Citation Counts in Evolution and Human Behavior, 1979–2008. *Evolutionary Psychology* **7**, 348–362.

WEBSTER, R. (2003). Let's re-write the scientific paper. *European Journal of Soil Science* **54**, 215–218.

WEINBERG, S. (1993). *Dreams of a Final Theory: The Scientist's Search for the Ultimate Laws of Nature*. Vintage books, New York.

WEINBERG, S. (2003). Four golden lessons. *Nature* **426**, 389.

WICKHAM, H. (2009). *ggplot: Elegant Graphics for Data Analysis*. Use R. Springer.

WOODS, G. (2002). *Research papers for dummies*. –For dummies. Hungry Minds.

WUCHTY, S., JONES, B. F., and UZZI, B. (2007). The Increasing Dominance of Teams in Production of Knowledge. *Science* **316**, 1036–1039.

ZINSSER, W. (2006). *On Writing Well*. Harper Collins Publishers.

Index

abstract, 103
academese, 96
Amazon, 21
American Physical Society, 36
article, types, 27
artwork, 77
author identifier, 21
authorship, 10, 34

bibliometrics, 13
  publication efficiency, 51
  services, 18
BioMed Central, 46
buzzwords, 108

Cham, 103, 106
Chomsky, 95
citation rate, 13
citing
  misconceptions, 83
  sources, 81, 83
code of honor, 36
collaboration
  annotation, 99
  editing, 96
Committee on publication ethics, 34
Content Management System, 46, 101
conventions, 5
COPE, 34, 72
Copernicus, 46, 48
copyright, 10, 43
Create-A-Research-Space, 139
Cressie, 133
cross-validation, 77
CrossRef, 24

Digital Object Identifier, 115
discussion, 138
DOAJ, 46
DOI, *see* Digital Object Identifier
DOI lookup, 84
double-blind review, 24
duplicate papers, 35

EASE guidelines, 65
editing, *see* collaboration
Elsevier, 44
empty verbs, 87
EndNote, 84
English
  however, 89
  On the use of I, 92
  spelling, 90
experiment, 4

figures, 77
first person, 92
flawed arguments, 76
Fox Hypothesis, 42
Free and Open Source Software, vii, 101
  LaTeX, 45
Free Culture, 43

Gerber, 125
ghost authors, 35
gift authors, 34
GNU General Public License, 100
golden rules, 139
Google Ad planner, 101
Google Scholar, 19
Google, removal tool, 82
graphics, 79
gray literature, 47

h-index, 14
Harper, 122

Harzing, 13, 14, 20
Hindawi, 46
how to be more specific?, 94
how to cite, 82
how to develop a style, 94
how to find a journal, 113
how to improve presentation, 76
however, 89
Hull, 72
hyperdisciplinarity, 9
hyperproduction, 34

images, 78
Impact Factor, 15
IMRaD, 71, 72
information production efficiency, 50
information resolution, 79
intellectual property, 10
inter-disciplinarity, 7
internet, vii
  growth rates, 48
introduction, 137
ISI Highly Cited, 29

JIF, *see* Journal's Impact Factor
journal list, 115
Journal's Impact Factor, 15
journals
  classification, 29
  list, 30

KISS, 138

Lessig, 43
line editing, 85
linking words, 89
logical moves, 139
logical reasoning, 74
Lulu, 46

maps, 79
Master journal list, 115
McBratney, 33, 56, 58, 61
McCandless, 80
MDPI, 46
Microsoft Academic Research, 19

nonsense paper, generator, 41

One Laptop Per Child, 8
one page summary, 105
Online Computer Library Center, 21
Open Access
  archiving, 49
  author-pays, 45
  publishing, 11
Open Access journals, 46
Open Peer Review, 38
Open Publishing, 45
Open Source Software, 100, 128
  R, 103
  list, 101

paragraphs, 85
Pebesma, 100, 120
peer review, 24
personal bias, 94
Phenomenon, movie, 92
photographs, 79
Pioneer plaque, 105
plagiarism, 36
pubcast, 120
Public Library of Science, 46
publication growth rates, 99
publication portals, 30
publishing
  commercial companies, 43
  efficiency, 51
  models, 11
  self-publishing, 47

RCI, *see* Relative Cost Index
recipe, 72
redundant words, 85, 110
references, import, 84
Relative Cost Index, 51
relative impact factor, 14
research article, 137
  assumption, 74
  classification, 26
  introduction, 7
  purpose, 12
  recipe, 72
  structure, 72
research quality, 13
ResearcherID, 22
rewriting, 110
Rossiter, 58, 74, 81

S–weave, 80, 99
Scholarpedia, 40
science
  classification, 7
  definition, 3
  problem areas, 34
  rules of science, 5
ScienceDirect, 30
scientific
  editing companies, 111
  information, 7, 9

publishing, 33
writing style, 95
scientists, definition, 6
SCImago, 114
SCOPUS, 19
ScribTex, 99
sections, 137
self-publishing, 34, 47
SHERPA, 30, 115
signposts, 140
software
Publish or Perish, 20
R, 103
Ruby on Rails, 46
Scientific WorkPlace, 81
spatio-temporal reference, 93
Springerlink, 30
Stallman, 100
statistical reasoning, 77
statistics
plots, 78
R book, 128
significance, 94
spatio-temporal reference, 93
stress curve, 125
subversion, 99
supergraphics, 80

ten rules, for getting published, 63
ten simple rules, 62
therefore, 89
thesis, 57
three Cs of research, 6
titles, 106
truth, 4
Tufte, 80
types of articles, 27

Vancouver guidelines, 72

Web of Knowledge, 18
Wikihow, 57
Wikimedia foundation, 98
WIKIPEDIA, 40, 97, 98
WorldCat, 21, 99
worst graphs, 78

Zinsser, 85, 86